ΑΛΕΞΑΝΔΡΕΙΟ Τ.Ε.Ι. ΘΕΣΣΑΛΟΝΙΚΗΣ
ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΚΩΝ ΕΦΑΡΜΟΓΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Πτυχιακή εργασία

« Συχνότητα και Κατανομή γραμμάτων στην Ελληνική Γλώσσα»

Των φοιτητών:

Παπαδόπουλου Κυριάκου
Αρ. Μητρώου: 02/1967

Τασιούδη Χρήστου
Αρ. Μητρώου:02/1971

Επιβλέπων καθηγητής

Βιολέττας Γεώργιος

Θεσσαλονίκη 2011

ΠΡΟΛΟΓΟΣ

Η παρούσα εργασία εκπονήθηκε κατά την διάρκεια του ακαδημαϊκού έτους 2010-2011 από τους φοιτητές του τμήματος Πληροφορικής, Παπαδόπουλο Κυριάκο και Τασιούδη Χρήστο, υπό την επίβλεψη του καθηγητή Βιολέττα Γεώργιου.

Αντικείμενο μελέτης αυτής της εργασίας είναι η συχνότητα και η κατανομή γραμμάτων στην ελληνική γλώσσα. Σκοπός της είναι να μελετηθεί η συχνότητα και κατανομή ελληνικών γραμμάτων όπως και ο βαθμός στον οποίο επηρεάζει η κρυπτογραφία την συχνότητα των γραμμάτων.

Ιδιαίτερη βαρύτητα δίνεται στην συχνότητα και κατανομή γραμμάτων και συμβόλων στους κωδικούς πρόσβασης που δημιουργούν οι Έλληνες χρήστες στα ηλεκτρονικά μέσα.

ΠΕΡΙΛΗΨΗ

Σε αυτή την πτυχιακή εργασία θα εξεταστεί η συχνότητα και η κατανομή γραμμάτων στην ελληνική γλώσσα. Αρχικά θα γίνει έρευνα στην βιβλιογραφία για την αγγλική γλώσσα και θα εξειδικευτεί στην ελληνική. Θα παρουσιαστεί λογισμικό που αναπτύχθηκε στα πλαίσια αυτής της εργασίας για την αναγνώριση της αρχικής γλώσσας προέλευσης ενός κρυπτοκειμένου και θα μελετηθεί επίσης ο βαθμός στον οποίο επηρεάζει η κρυπτογραφία την συχνότητα των γραμμάτων. Τέλος, θα παρουσιαστεί πρωτογενής έρευνα που διεξήχθη για την κατανομή των γραμμάτων στους κωδικούς πρόσβασης που δημιουργούν οι Έλληνες χρήστες στα ηλεκτρονικά μέσα.

ABSTRACT

In this thesis will be examined the frequency and distribution of letters in Greek written language. Initial literature research was done about the English language, and then, it was specialized for the Greek language. Software development will also be presented, to identify the original source language of a cipher text. In addition, the level in which cryptography can affect the letter frequency will be considered. Finally, primary research will be presented, conducted through the web over the distribution of letters on passwords that are created on electronic media, by Greek speaking web users.

Ευχαριστίες

Η παρούσα πτυχιακή εργασία δεν θα είχε ολοκληρωθεί με την σημερινή της μορφή χωρίς την βοήθεια και την στήριξη κάποιων ανθρώπων που θα θέλαμε να ευχαριστήσουμε θερμά.

Κατ' αρχήν θέλουμε να εκφράσουμε τις θερμές μας ευχαριστίες στον καθηγητή κ. Γεώργιο Βιολέττα, επιβλέποντα της πτυχιακής εργασίας, του οποίου οι συμβουλές και συμπαράσταση ήταν πολύτιμες καθ' όλη την ολοκλήρωσης της εργασίας.

Επίσης θα θέλαμε να ευχαριστήσουμε τους φίλους και συμφοιτητές μας, Πλιάκα Αχιλλέα και Τσέκο Κωνσταντίνο, των οποίων οι προτροπές και ιδέες επηρέασαν τον τρόπο σκέψης μας και συνέβαλαν στην βελτίωση της σημερινής μορφής της εργασίας μας.

Παπαδόπουλος Κυριάκος,
Τασιούδης Χρήστος,

Θεσσαλονίκη 2011.

ΕΥΡΕΤΗΡΙΟ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΚΕΦΑΛΑΙΟ 2

ΚΕΦΑΛΑΙΟ 3

ΚΕΦΑΛΑΙΟ 4

ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ

ΕΥΡΕΤΗΡΙΟ ΔΙΑΓΡΑΜΜΑΤΩΝ

ΕΥΡΕΤΗΡΙΟ ΕΙΚΟΝΩΝ

ΕΙΣΑΓΩΓΗ

Το αντικείμενο αυτής της πτυχιακής είναι η μελέτη της συχνότητας και κατανομής των γραμμάτων στην ελληνική γλώσσα. Σκοπός αυτής της εργασίας είναι να μελετηθεί στη βιβλιογραφία η συχνότητα και η κατανομή των ελληνικών γραμμάτων και με την βοήθεια ανάπτυξης λογισμικού να γίνει προσπάθεια για την ανίχνευση της γλώσσας προέλευσης ενός κρυπτοκειμένου, αλλά και την μελέτη της επιρροής της κρυπτογραφίας στην συχνότητα και κατανομή των γραμμάτων. Επίσης, ερευνάται η συχνότητα και κατανομή των γραμμάτων σε κωδικούς πρόσβασης από Έλληνες χρήστες σε ηλεκτρονικά μέσα. Η παρουσίαση των επιμέρους θεμάτων και αποτελεσμάτων της πτυχιακής εργασίας οργανώνεται ως εξής:

Στο κεφάλαιο 1 της πτυχιακής παρουσιάζονται στατιστικά στοιχεία που μελετήθηκαν στην διεθνή βιβλιογραφία για την συχνότητα και την κατανομή των γραμμάτων στην αγγλική γλώσσα. Έπειτα η μελέτη αυτή εξειδικεύεται στην ελληνική γλώσσα. Με την βοήθεια του Εθνικού Θησαυρού της Ελληνικής Γλώσσας, μελετώνται παράμετροι όπως οι συχνότερες ελληνικές λέξεις, τα συχνότερα ελληνικά γράμματα και το μήκος λέξης σε διαφορετικά κειμενικά μέσα. Τέλος, εξετάζεται το παιχνίδι «scrabble», το οποίο χρησιμοποιεί την συχνότητα των γραμμάτων.

Στο κεφάλαιο 2 γίνεται αναφορά στην κρυπτολογία η οποία χωρίζεται στην κρυπτογραφία και την κρυπτανάλυση. Δίνονται οι βασικοί ορισμοί των οντοτήτων των εννοιών αυτών και παρουσιάζονται τα είδη κρυπτοσυστημάτων και αλγόριθμοι κρυπτογράφησης. Από τους σημαντικότερους αλγόριθμους κρυπτογράφησης είναι ο Data Encryption Standard (DES), καθώς και βελτιώσεις του οι οποίες αναλύονται στο κεφάλαιο αυτό.

Στο κεφάλαιο 3 παρουσιάζονται δύο εφαρμογές λογισμικού που αναπτύχθηκαν στα πλαίσια αυτής της εργασίας. Η πρώτη εφαρμογή που αναπτύχθηκε αφορά την δημιουργία ενός κρυπτοκειμένου ώστε μέσα από αυτό να μπορέσει να ανιχνευθεί η αρχική γλώσσα προέλευσης του κρυπτοκειμένου. Η δεύτερη εφαρμογή, αποτελεί και την ολοκληρωμένη διαδικασία της κρυπτογράφησης. Ένα αρχικό κείμενο κρυπτογραφείται και αποκρυπτογραφείται. Σημειώνεται επίσης ότι επιπλέον δυνατότητα της εφαρμογής αυτής είναι η μέτρηση των χαρακτήρων στο αρχικό και στο κρυπτογραφημένο κείμενο. Από αυτή τη μέτρηση προκύπτει το συμπέρασμα

της επιρροής της κρυπτογραφίας στην συχνότητα των γραμμάτων. Επίσης γίνονται συγκρίσεις των συχνοτήτων στο αρχικό κείμενο μέσω της υπάρχουσας βιβλιογραφίας.

Στο κεφάλαιο 4 γίνεται αναφορά στους κωδικούς πρόσβασης στα ηλεκτρονικά μέσα. Δίνονται ορισμοί και περιγράφονται οι ισχυροί και αδύναμοι κωδικοί. Επίσης αναλύονται τεχνικές και κατευθύνσεις για την δημιουργία ενός ισχυρού κωδικού πρόσβασης.

Στο κεφάλαιο 5 παρουσιάζεται έρευνα που διεξήχθη στα πλαίσια της παρούσας εργασίας. Η έρευνα διεξήχθη με την χρήση ερωτηματολογίου όπου το δείγμα αποκάλυψε τεχνικές που χρησιμοποιεί στους δικούς του κωδικούς, όπως το μέγεθος του κωδικού, η πολυπλοκότητα και τρόποι αποθήκευσής του. Τέλος παρουσιάζονται οι 100 πιο αδύναμοι κωδικοί πρόσβασης που δεν πρέπει να χρησιμοποιούν οι Έλληνες χρήστες στα ηλεκτρονικά μέσα.

ΚΕΦΑΛΑΙΟ 1

1. ΣΥΧΝΟΤΗΤΑ ΚΑΙ ΚΑΤΑΝΟΜΗ ΓΡΑΜΜΑΤΩΝ ΣΤΗΝ ΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ

1.1. ΕΙΣΑΓΩΓΗ

Σε αυτό το κεφάλαιο εξετάζεται η συχνότητα και η κατανομή των γραμμάτων στον γραπτό λόγο, σε διάφορες γλώσσες και εξειδικεύεται στην ελληνική. Η μελέτη αρχίζει με την αναζήτηση της συχνότητας και της κατανομής των γραμμάτων σε ξένες γλώσσες με ιδιαίτερη αναφορά στην αγγλική. Η εξειδίκευση της μελέτης θα γίνει για την ελληνική καθώς έως σήμερα δεν υπάρχουν δημοσιεύσεις σχετικά με την συχνότητα εμφάνισης των γραμμάτων σε κείμενα και βιβλία για την ελληνική γλώσσα. Επίσης, γίνεται έρευνα σχετικά με τις συχνότερες λέξεις που χρησιμοποιούνται στην ελληνική γλώσσα, αλλά και την σχέση μεταξύ του είδους κειμένου και το ποσοστό λέξεων και γραμμάτων.

1.2. ΟΡΙΣΜΟΙ

Ξεκινώντας θεωρείται σκόπιμο να δοθούν οι ορισμοί της συχνότητας και της κατανομής. Γενικά ως **συχνότητα** θεωρείται ο ρυθμός επανάληψης ενός φαινομένου. Η **κατανομή** θεωρείται ότι κατά κάποιο τρόπο μοιράζει ή διανέμει αντικείμενα με βάση κάποιες συγκεκριμένες ιδιότητες που τα καθιστούν διακριτά από άλλα. Συγκεκριμένα στην ανάλυση γλώσσας, οι περισσότερες γλώσσες παρουσιάζουν στη δομή τους (γράμματα ή συνδυασμούς γραμμάτων) κάποια ορισμένη κατανομή με μέγιστα και ελάχιστα, τα οποία μπορούν να χαρακτηρίσουν τη γλώσσα αυτή. Με τον υπολογισμό της κατανομής των γραμμάτων μέσα στη γλώσσα βρίσκεται ένα μέτρο που το ακολουθούν όλα τα κείμενα της γλώσσας αυτής. Για παράδειγμα, όπως θα επεξηγηθεί παρακάτω, για την αγγλική γλώσσα το E τείνει να είναι το πιο κοινό γράμμα (με τις περισσότερες επαναλήψεις σε ένα οποιοδήποτε κείμενο) ενώ το Z τείνει να είναι το πιο σπάνια συναντούμενο γράμμα.

Πρέπει να τονιστεί ότι για κάποιες γλώσσες υπάρχει επίσημη βιβλιογραφία (Beutelspacher, 2005; Pratt, 1942; Singh, 1999) που αναφέρει την συχνότητα και την κατανομή των γραμμάτων τους, ενώ έως σήμερα ελάχιστη έρευνα έχει γίνει για την ελληνική γλώσσα στον αντίστοιχο τομέα. Παρακάτω θα γίνει αναφορά για την

ανάλυση της αγγλικής γλώσσας, η οποία παρουσιάζεται με στατιστικά δεδομένα, ενώ, έπειτα, μέσα από έρευνα, θα εξεταστούν τα στατιστικά δεδομένα που υπάρχουν για την ελληνική γλώσσα.

1.3. ΜΕΛΕΤΗ ΣΥΧΝΟΤΗΤΗΤΩΝ ΣΤΗΝ ΑΓΓΛΙΚΗ ΓΛΩΣΣΑ

Για την ανάλυση της αγγλικής γλώσσας ως προς την συχνότητα και την κατανομή των γραμμάτων της, κατά καιρούς έχουν δημοσιευτεί βιβλία και άρθρα που αναφέρουν σημαντικές παραμέτρους όπως:

1. η συχνότητα αγγλικών γραμμάτων,
2. η συχνότητα αγγλικών γραμμάτων ως πρώτο σε μία λέξη,
3. οι πιο συχνές λέξεις στην αγγλική γλώσσα,
4. τα πιο συχνά μέρη του λόγου (ρήμα, ουσιαστικό, επίθετο, πρόθεση),

τις οποίες και θα αναλυθούν παρακάτω.

1.3.1. Συχνότητα αγγλικών γραμμάτων.

Σύμφωνα με τους Beker και Piper οι οποίοι το 1982 δημοσίευσαν το βιβλίο «Cipher systems: the protection of communications» (Beker & Piper, 1982), παρατίθεται ο παρακάτω πίνακας ο οποίος παρέχει την συχνότητα των πιο συχνών γραμμάτων που χρησιμοποιούνται στην αγγλική γλώσσα.

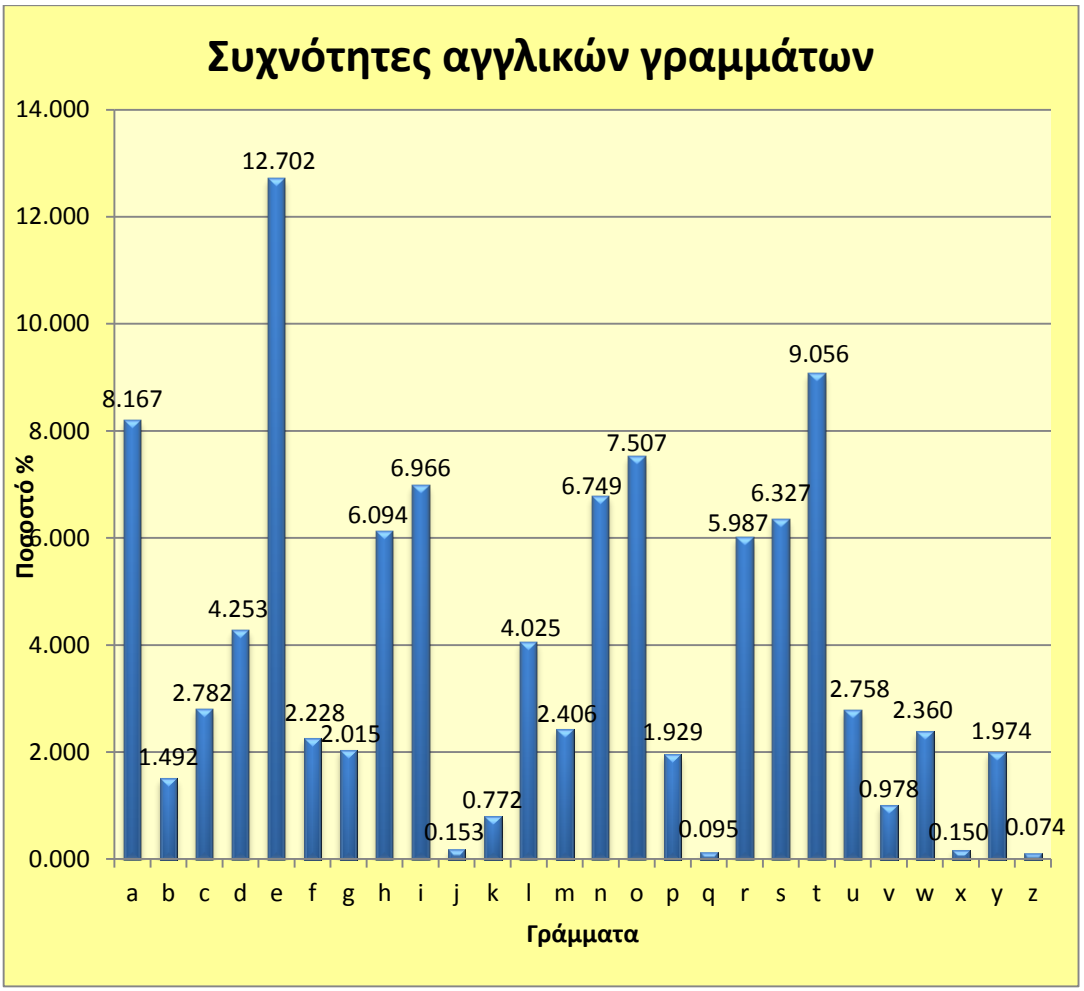
Πίνακας 2.1: Συχνότητα αγγλικών γραμμάτων.

Συχνότητα αγγλικών γραμμάτων			
Γράμμα	**Συχνότητα**	**Γράμμα**	**Συχνότητα**
a	8.167%	n	6.749%
b	1.492%	o	7.507%
c	2.782%	p	1.929%
d	4.253%	q	0.095%
e	12.702%	r	5.987%
f	2.228%	s	6.327%
g	2.015%	t	9.056%
h	6.094%	u	2.758%
i	6.966%	v	0.978%

Γράμμα	Συχνότητα	Γράμμα	Συχνότητα
j	0.153%	w	2.360%
k	0.772%	x	0.150%
l	4.025%	Y	1.974%
m	2.406%	z	0.074%

Συμπεράσματα: Όπως συμπεραίνεται από τον παραπάνω πίνακα, το πιο συχνό γράμμα που χρησιμοποιείται στην αγγλική γλώσσα είναι το «e» με ποσοστό εμφάνισής του 12,7 % , ακολουθεί το γράμμα «t» με ποσοστό 9 % και τρίτο στην κατάταξη ανέρχεται το «a» με ποσοστό εμφάνισης 8,1 %. Στα παρακάτω διαγράμματα παρουσιάζονται τα ποσοστά εμφάνισης των αγγλικών γραμμάτων αλφαβητικά.

Στο παρακάτω διάγραμμα παρουσιάζονται οι συχνότητες των αγγλικών γραμμάτων.



Διάγραμμα 2.1: Συχνότητες αγγλικών γραμμάτων σύμφωνα με τους Beker και Piper.

Αυτή η έρευνα επαληθεύεται και από την επίσημη ιστοσελίδα του Oxford (‘Oxford Dictionaries Online’, 2011) που παρέχει λεξιλόγιο της αγγλικής γλώσσας.

1.3.2. Συχνότητα αγγλικών γραμμάτων ως πρώτα σε λέξεις.

Μία άλλη παράμετρος η οποία εξετάζεται σχετικά με την ανάλυση κατανομής και συχνότητας των αγγλικών γραμμάτων, σύμφωνα με όσα παρέχουν δημοσιευμένα άρθρα και βιβλία (Rayner & Duffy, 1986; Shannon, 1951), είναι η συχνότητα των γραμμάτων όταν αυτά αποτελούν το πρώτο γράμμα μιας λέξης.

Σύμφωνα με το Oxford English Corpus το οποίο διεξήγε έρευνα σχετικά με την κατανομή γραμμάτων, σε συνεργασία με το Oxford English Dictionary (‘Oxford Dictionaries Online’, 2011), στην επίσημη ιστοσελίδα παραθέτει αποτελέσματα αναφορικά με τα γράμματα που εμφανίζονται πρώτα σε μία λέξη. Βασισμένοι σε

υπολογισμούς από τον Mandani μπορούν να παρουσιαστούν τα συχνότερα αγγλικά γράμματα που εμφανίζονται πρώτα σε μία λέξη (Madnani, 2007) τα οποία και αναφέρονται στον πίνακα 1.2.

Πίνακας 2.2: Συχνότερα αγγλικά γράμματα ως πρώτα σε μία λέξη.

Γράμμα	Συχνότητα	Γράμμα	Συχνότητα
a	11.602%	n	2.365%
b	4.702%	o	6.264%
c	3.511%	p	2.545%
d	2.670%	q	0.173%
e	2.000%	r	1.653%
f	3.779%	s	7.755%
g	1.950%	t	16.671%
h	7.232%	u	1.487%
i	6.286%	v	0.619%
j	0.631%	w	6.661%
k	0.690%	x	0.005%
l	2.705%	y	1.620%
m	4.374%	z	0.050%

1.3.3. Συχνότερες αγγλικές λέξεις.

Σε μία από τις πρώτες δημοσιεύσεις πάνω στην ανάλυση της αγγλικής γλώσσας (Johansson & Hofland, 1989), παρουσιάστηκαν οι πιο συχνές λέξεις όπως φαίνεται στον παρακάτω πίνακα:

Πίνακας 2.2: Συχνότερες αγγλικές λέξεις το 1989.

α/α	Λέξη	Συχνότητα
1	the	68315
2	of	35716
3	and	27856
4	to	26760

α/α	Λέξη	Συχνότητα
5	a	22744
6	in	21108
7	that	11188
8	is	10978
9	was	10499
10	it	10010
11	for	9299
12	he	8776
13	as	7337
14	with	7197
15	be	7186
16	on	7027
17	I	6696
18	his	6266

Πρόσφατα όμως, όπως ανακοίνωσε το Oxford Dictionary (‘Oxford Dictionaries Online’, 2011) παρατηρήθηκαν κάποιες διαφορές σε σχέση με τα αποτελέσματα που προέκυψαν από την έρευνα του Johansson το 1989. Για παράδειγμα η λέξη «be» από την 15η θέση του πίνακα του Johansson, το Oxford μετατοπίζει την συγκεκριμένη λέξη στην 2η θέση. Προφανώς οι αλλαγές οφείλονται και στο πέρασμα του χρόνου, στην αλλαγή του τρόπου γραφής της αγγλικής αλλά το κυριότερο στο μεγάλο όγκο δεδομένων που βασίζεται το Oxford. Τα αποτελέσματα παρουσιάζονται στον πίνακα 1.4 παρακάτω και είναι βασισμένα σε εφημερίδες, άρθρα και λογοτεχνικά διηγήματα. Αξίζει να σημειωθεί ότι η έρευνα βασίζεται σε πάνω από 1 δισεκατομμύριο λέξεις. Στον πίνακα 1.4 ενδεικτικά παρουσιάζονται οι 20 πιο συχνές αγγλικές λέξεις, βασισμένες στο Oxford English Dictionary (‘Oxford Dictionaries Online’, 2011).

Πίνακας 2.3: Συχνότερες αγγλικές λέξεις βασισμένες στο Oxford English Dictionary

α/α	Λέξη	α/α	Λέξη
1	the	11	it
2	be	12	for

α/α	Λέξη	α/α	Λέξη
3	to	13	not
4	of	14	on
5	and	15	with
6	a	16	he
7	in	17	as
8	that	18	you
9	have	19	do
10	I	20	at

1.4. ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ ΓΙΑ ΤΗΝ ΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ.

«Η ποσοτική διερεύνηση της γλωσσικής δομής μιας γλώσσας αποτελεί έναν από τους σημαντικότερους στόχους της σύγχρονης γλωσσολογικής έρευνας. Η παροχή Ηλεκτρονικών Σωμάτων Κειμένων (ΗΣΚ) σε συνδυασμό με την ανάπτυξη κατάλληλών υπολογιστικών και στατιστικών τεχνικών για τη διαχείριση γλωσσικών δεδομένων έχουν εισαγάγει ποσοτικές μεθόδους στην ανάλυση του συνόλου των γλωσσολογικών επιπέδων» (BOD & Jennifer, 2003). Η αυξημένη χρήση ποσοτικών μεθόδων στη γλωσσική ανάλυση την τελευταία δεκαετία εξηγείται εν μέρει και από την αυξημένη αποτελεσματικότητα που επιδεικνύουν αυτές σε πλήθος εργασιών που σχετίζονται με την Επεξεργασία Φυσικής Γλώσσας (Manning, Schütze, & MITCogNet, 1999).

Η γενικότερη αυτή τάση στη γλωσσολογική έρευνα επιβεβαιώνεται και στην μελέτη της Νέας Ελληνικής (ΝΕ) γλώσσας. Σε μια πρόσφατη έρευνα (Mikros υπό δημοσίευση) διαπιστώθηκε ότι η χρήση ποσοτικών μεθόδων στις γλωσσολογικές έρευνες της ΝΕ ακολουθεί εκθετική αύξηση, με το ποσοστό της να έχει πενταπλασιαστεί την δεκαετία του 90 σε σχέση με την δεκαετία του 80.

Μια σημαντική εξέλιξη στην αξιοποίηση των ποσοτικών μεθόδων στην ελληνική γλωσσολογία αποτέλεσε και η ανάπτυξη του Εθνικού Θησαυρού της Ελληνικής Γλώσσας (ΕΘΕΓ) (Goutsos, 2010; Hatzigeorgiu, et. al., 2000) από το Ινστιτούτο Επεξεργασίας του Λόγου.

1.4.1. Ποσοτική γλωσσολογία.

Ποσοτική Γλωσσολογία (ΠΓ) είναι ο κλάδος εκείνος της Γλωσσολογίας που ασχολείται με την ποσοτική ανάλυση της γλωσσικής δομής και τη γλωσσολογική ερμηνεία της (Mikros, 2002).
Η ποσοτική ανάλυση χρησιμοποιείται για να ολοκληρωθεί η ποιοτική ανάλυση που διεξάγει ο κλάδος της θεωρητικής γλωσσολογίας. Γενικότερα, η χρήση ποσοτικών μεθόδων όπως αυτές που θα περιγραφούν παρακάτω λειτουργούν συμπληρωματικά με τις ποιοτικές θεωρήσεις ως προς την κατανόηση του γλωσσικού φαινομένου.

1.4.2. Ηλεκτρονικά Σώματα Κειμένων.

Η ποσοτική αντιμετώπιση της γλωσσικής χρήσης θα ήταν αδύνατη εάν δεν υπήρχαν τα Ηλεκτρονικά Σώματα Κειμένων. Υπάρχουν διάφοροι ορισμοί που μπορούν να δοθούν για να περιγράψουν ένα Ηλεκτρονικό Σώμα Κειμένου, οι σημαντικότεροι εκ των οποίων είναι:

1. Σύμφωνα με τους McEnery και Wilson (McEnery, Wilson, et. al. 1996) το Ηλεκτρονικό Σώμα Κειμένων στην σύγχρονη γλωσσολογία μπορεί να περιγραφεί ως ένα σώμα με πεπερασμένο αριθμό κειμένων σε ηλεκτρονική μορφή, τα οποία έχουν επιλεγεί με τέτοιο τρόπο, ώστε να αποτελούν όσο το δυνατόν πιο αντιπροσωπευτικά δείγματα της γλωσσικής ποικιλίας, που μελετάται.
2. Είναι η συλλογή τμημάτων γλώσσας τα οποία επιλέγονται και διατάσσονται σύμφωνα με συγκεκριμένα γλωσσολογικά κριτήρια έτσι ώστε να χρησιμοποιηθούν ως αντιπροσωπευτικό δείγμα μιας συγκεκριμένης γλώσσας (Leech et. al, 1996). Δηλαδή είναι μία συλλογή κειμένων η οποία είναι κωδικοποιημένη για τυποποιημένες (standardized) και ομοιογενείς εργασίες ανάκτησης γλωσσικής πληροφορίας.

1.4.3. Εθνικός Θησαυρός Ελληνικής Γλώσσας.

Ο Εθνικός Θησαυρός της Ελληνικής Γλώσσας (ΕΘΕΓ), τα εργαλεία για την κατασκευή του, καθώς και όλα τα εργαλεία που χρησιμοποιήθηκαν για τα

αποτελέσματα που έχει εξάγει, έχουν κατασκευαστεί από το Ινστιτούτο Επεξεργασίας του Λόγου.

Ο ΕΘΕΓ είναι ένα δυναμικό ΗΣΚ της γραπτής Νέας Ελληνικής γλώσσας. Αυτή τη στιγμή περιλαμβάνει περισσότερα από 48.000 κείμενα που δημοσιεύτηκαν μετά το 1976 και αποτελούνται από περισσότερες από 33 εκατομμύρια λέξεις. Για τον ΕΘΕΓ έχει κατασκευαστεί μια διεπαφή χρήστη (web interface) για το διαδίκτυο (βλ. 'ΕΘΕΓ - Εθνικός Θησαυρός Ελληνικής Γλώσσας', 2011) και μπορεί να χρησιμοποιηθεί με συνδρομή από οποιονδήποτε ενδιαφερόμενο ερευνητή.

Τα κείμενα που περιέχει ο ΕΘΕΓ έχουν κατηγοριοποιηθεί με βάση το σύστημα PAROLE ('PAROLE', 2011), το οποίο ακολουθεί τις οδηγίες TEI (Sperberg-McQueen & Burnard, 1995) και EAGLES ('EAGLES', 1994). Πριν την εισαγωγή τους τα κείμενα κατηγοριοποιούνται με βάση το Μέσο Δημοσίευσης (Medium), το Γένος (Genre), το Θέμα (Topic), το Ειδικότερο Γένος (Detailed Genre), το Ειδικότερο Θέμα (Detailed Topic), τον Εκδότη, τον Συγγραφέα και την Ημερομηνία Έκδοσης. Με βάση το Μέσο Δημοσίευσης τα κείμενα κατατάσσονται σε τέσσερις κατηγορίες οι οποίες φαίνονται στον Πίνακα 1.5, όπου έχει σημειωθεί και η αναλογία του αριθμού των λέξεων, όπως αυτή έχει διαμορφωθεί σήμερα.

Πίνακας 2.4: Κατανομή κειμένων στον ΕΘΕΓ.

Μέσο Δημοσίευσης	Ποσοστό λέξεων στον ΕΘΕΓ
Βιβλίο	9,41%
Εφημερίδα	61,29%
Περιοδικό	5,89%
Αδιευκρίνιστο	23,08%

Σημειώνεται ότι μια σημαντική ώθηση δόθηκε όταν αριθμός εκδοτών δέχθηκε με προθυμία να παραχωρήσει τα απαραίτητα πνευματικά δικαιώματα για μια τέτοια εργασία. Η κάποια ασυμμετρία που παρουσιάζει η κατανομή των διάφορων κατηγοριών του ΕΘΕΓ σήμερα, οφείλεται στην μη σύμμετρη εμπλοκή μεγαλύτερου αριθμού εκδοτών.

Στους παρακάτω πίνακες (1.6, 1.7) παρουσιάζονται τα επίσημα αποτελέσματα που υπάρχουν στην ιστοσελίδα του ΕΘΕΓ (βλ. 'ΕΘΕΓ - Εθνικός Θησαυρός

Ελληνικής Γλώσσας', 2011). Ο πίνακας 1.6 παρουσιάζει τις **100 συχνότερες ελληνικές λέξεις**, ενώ ο πίνακας 1.7 τα **100 συχνότερα ελληνικά λήμματα.**
Πρώτα όμως θεωρείται αναγκαίο να αναφερθούν κάποιες πληροφορίες σχετικά με τις πηγές των κειμένων που βασίζονται αυτά τα αποτελέσματα. Πρέπει να σημειωθεί ότι:

1. Το σώμα κειμένων του Ινστιτούτου Επεξεργασίας του Λόγου (ΙΕΛ) αναπτύχθηκε επί σειρά ετών και σήμερα περιλαμβάνει περισσότερες από 47.000.000 λέξεις, ενώ εμπλουτίζεται συνεχώς. Όλα τα κείμενα του ΕΘΕΓ είναι επιλεγμένα, έτσι ώστε να αντικατοπτρίζουν την πραγματική εικόνα της σύγχρονης γλώσσας.
2. Το Σώμα Κειμένων του ΙΕΛ περιλαμβάνει αποκλειστικά δείγματα γραπτού λόγου. Προφορικός λόγος δεν έχει περιληφθεί στην παρούσα έκδοση του Σώματος.
3. Τα κείμενα που περιλαμβάνονται στο Σώμα Κειμένων του ΙΕΛ έχουν επιλεγεί ως αντιπροσωπευτικά της σύγχρονης Ελληνικής γλώσσας και χρονολογούνται, στην πλειονότητά τους, από το 1990 και μετά. Αποφεύγονται τα κείμενα με διαλεκτικές ή άλλες ιδιαιτερότητες και προτιμώνται κείμενα με υψηλή αναγνωσιμότητα (εφημερίδες μεγάλης κυκλοφορίας, βιβλία με υψηλές πωλήσεις κτλ).
4. Με στόχο την αντιπροσώπευση διαφορετικών επιπέδων λόγου, επιλέχθηκαν κείμενα από πολλές πηγές , που καλύπτουν ποικίλα κειμενικά είδη με ποικίλη θεματολογία.
5. Τα κείμενα αυτά έχουν παραχωρηθεί νόμιμα στο ΙΕΛ από τους κατόχους τους και διατίθενται μόνο για ερευνητικούς σκοπούς.

Πίνακας 2.5: : Οι 100 συχνότερες ελληνικές λέξεις.

Οι 100 συχνότερες ελληνικές λέξεις					
α/α	**Λέξη**	**Συχνότητα(%)**	**α/α**	**Λέξη**	**Συχνότητα(%)**
1	και	34,1810	**51**	είχε	1,5245
2	του	23,2783	**52**	μου	1,3739
3	το	22,6643	**53**	μπορεί	1,3705
4	να	20,8845	**54**	πολύ	1,3422
5	της	19,6292	**55**	τι	1,3218

α/α	Λέξη	Συχνότητα(%)	α/α	Λέξη	Συχνότητα(%)
6	η	16,8816	**56**	όταν	1,3059
7	την	16,4742	**57**	προς	1,2790
8	που	14,5250	**58**	μόνο	1,2643
9	με	13,7311	**59**	μετά	1,2444
10	από	12,6343	**60**	γιατί	1,2310
11	ο	12,4455	**61**	ενώ	1,1901
12	για	12,4307	**62**	σας	1,1610
13	των	11,2146	**63**	στους	1,1495
14	τα	10,0034	**64**	μέσα	1,0849
15	είναι	9,3521	**65**	όχι	1,0547
16	θα	8,5768	**66**	αυτά	1,0546
17	οι	8,4911	**67**	κάθε	1,0180
18	δεν	8,4014	**68**	σήμερα	0,9779
19	στο	7,9312	**69**	υπάρχει	0,9713
20	σε	7,9257	**70**	μία	0,9633
21	ότι	7,5451	**71**	πιο	0,9454
22	τη	7,4360	**72**	σ	0,9120
23	στην	7,3065	**73**	χρόνια	0,9037
24	τον	7,1236	**74**	έτσι	0,8904
25	τους	6,9081	**75**	όλα	0,8768
26	τις	6,0330	**76**	θέμα	0,8625
27	στη	3,7821	**77**	μέχρι	0,8384
28	έχει	3,2118	**78**	τώρα	0,8374
29	μια	3,1003	**79**	οποίο	0,8337
30	ένα	3,0883	**80**	ακόμη	0,8304
31	αυτό	3,0037	**81**	χωρίς	0,8095
32	αλλά	2,8517	**82**	ελλάδα	0,8025
33	στα	2,6485	**83**	κυβέρνηση	0,8001
34	στις	2,6192	**84**	1	0,7907
35	μας	2,5841	**85**	χθες	0,7792
36	ή	2,5695	**86**	πως	0,7789

α/α	Λέξη	Συχνότητα(%)	α/α	Λέξη	Συχνότητα(%)
37	ήταν	2,4180	**87**	μεταξύ	0,7622
38	στον	2,3129	**88**	όπου	0,7570
39	κ	2,2758	**89**	είπε	0,7487
40	αν	2,1351	**90**	πολιτική	0,7376
41	όπως	2,1287	**91**	καθώς	0,7324
42	ως	1,9530	**92**	ένας	0,7258
43	κατά	1,9293	**93**	απ	0,7157
44	όμως	1,9030	**94**	αυτές	0,7147
45	αυτή	1,8118	**95**	επίσης	0,7064
46	έχουν	1,7810	**96**	2	0,7061
47	πρέπει	1,7590	**97**	δηλαδή	0,7048
48	οποία	1,5515	**98**	πριν	0,6933
49	δύο	1,5412	**99**	ούτε	0,6889
50	κι	1,5373	100	σύμφωνα	0,6886
Σύνολο: 437,9623‰					

Πίνακας 2.6: Τα 100 συχνότερα λήμματα στην ελληνική γλώσσα.

Τα 100 πιο συχνά λήμματα στην ελληνική γλώσσα					
α/α	**Λήμμα**	**Εμφανίσεις**	**α/α**	**Λήμμα**	**Εμφανίσεις**
1	ο	7926545	**51**	χώρα	69479
2	εγώ	7376916	**52**	ίδιος	68834
3	μου	2595432	**53**	μέσος	67931
4	και	1679098	**54**	μόνο	67574
5	στου	1305211	**55**	όλο	67545
6	το	1065387	**56**	όσος	65902
7	να	982003	**57**	κύριος	65626
8	η	793461	**58**	πολιτική	64443
9	που	682811	**59**	καλός	64377
10	με	645489	**60**	θέμα	64271
11	από	629435	**61**	υπουργός	63627

α/α	Λέξη	Συχνότητα(%)	α/α	Λέξη	Συχνότητα(%)
12	για	607089	**62**	ακόμα	62394
13	είμαι	589317	**63**	τι	62141
14	τα	470215	**64**	βρίσκω	61728
15	ένας	448874	**65**	όταν	61390
16	αυτός	423736	**66**	μέσο	61263
17	δεν	414620	**67**	προς	60124
18	έχω	412737	**68**	μετά	59181
19	θα	403196	**69**	θέλω	58073
20	σε	372576	**70**	κυβέρνηση	57999
21	ότι	354677	**71**	γιατί	57875
22	τη	349538	**72**	μην	57408
23	τους	324749	**73**	ελληνικός	57249
24	οποίος	185208	**74**	ενώ	55941
25	πολύς	154281	**75**	πρόεδρος	53083
26	λέγω	149607	**76**	κάποιος	52308
27	όλος	135368	**77**	μέσα	52118
28	αλλά	134757	**78**	μήνας	51723
29	άλλος	132227	**79**	κανένας	50136
30	μπορώ	131952	**80**	όχι	49585
31	ή	120796	**81**	βλέπω	47985
32	αν	116817	**82**	κάθε	47857
33	γίνομαι	109684	**83**	ελλάδα	47161
34	κ	106990	**84**	λόγος	47090
35	κατά	105641	**85**	έργο	46678
36	υπάρχω	105393	**86**	πρόβλημα	46369
37	όπως	100074	**87**	θέση	46075
38	κάνω	98753	**88**	σήμερα	45974
39	μεγάλος	96038	**89**	δημόσιος	45386
40	πρέπει	94466	**90**	εθνικός	45050
41	νέος	93145	**91**	παίρνω	44647
42	ως	91808	**92**	λίγος	44545

α/α	Λέξη	Συχνότητα(%)	α/α	Λέξη	Συχνότητα(%)
43	όμως	89465	**93**	πιο	44456
44	πολιτικός	86276	**94**	οικονομικός	44429
45	δύο	83599	**95**	ελλάς	44416
46	πολύ	80616	**96**	ευρωπαϊκός	43302
47	δίνω	79391	**97**	δίδω	43018
48	μόνος	79046	**98**	σ	42875
49	χρόνος	78255	**99**	χρόνιος	42839
50	πρώτος	76388	**100**	έτσι	41861
Σύνολο: 36.482.494					

1.5. ΒΑΣΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΤΩΝ 1000 ΣΥΧΝΟΤΕΡΩΝ ΛΕΞΕΩΝ ΚΑΙ ΛΗΜΜΑΤΩΝ

1.5.1. Σύγκριση των 1000 συχνότερων λέξεων στις δύο εκδόσεις του ΕΘΕΓ.

Στο (Hatzigeorgiu, Mikros, & Carayannis, 2001) δημοσιεύεται η πρώτη λίστα των 100 πιο συχνών λέξεων και λημμάτων της ΝΕ αξιοποιώντας την πρώτη έκδοση του ΕΘΕΓ, η οποία αριθμούσε συνολικά 13 εκ. λέξεις. Η αύξηση του μεγέθους του ΕΘΕΓ στα 33 εκ. λέξεις κατέστησε την επανεκτίμηση των συχνότερων λέξεων της ΝΕ επιτακτική.

Οι 1000 συχνότερες λέξεις στον ΕΘΕΓ των 13 εκ. λέξεων αποτελούν το 59,9% του συνόλου των λέξεων, ενώ στον ΕΘΕΓ των 33 εκ. λέξεων το 60,4%. Παρατηρείται επομένως ότι σε απόλυτους αριθμούς οι 1000 συχνότερες λέξεις καταλαμβάνουν σχετικά σταθερό ποσοστό του συνολικού λεξιλογίου ενός ΗΣΚ, ακόμα και όταν προστίθεται σημαντικός αριθμός νέων κειμένων και το μέγεθος του γίνεται 2,5 φορές μεγαλύτερο.

Για να εξεταστεί με περισσότερη λεπτομέρεια το μέγεθος της διαφοροποίησης των δύο εκδόσεων του ΕΘΕΓ ως προς το φάσμα των συχνών λέξεων έγινε η καταμέτρηση της συχνότητας των 1000 πιο συχνών λέξεων και στις δύο εκδόσεις του ΕΘΕΓ. Οι λίστες συχνότητας που δημιουργήθηκαν εξισώθηκαν ως προς τις

λέξεις που περιείχαν και για κάθε λέξη καταγράφηκε η συχνότητα της στις δύο εκδόσεις του ΕΘΕΓ (13 εκ. και 33 εκ.), η σχετική συχνότητα εμφάνισής της (ποσοστό της απόλυτης συχνότητας χρήσης ως προς το συνολικό μέγεθος του ΕΘΕΓ) και η σχετική θέση της λέξης στη λίστα (κατάταξη) για κάθε έκδοση. Εν συνεχεία υπολογίστηκαν η διαφορά συχνότητας και σχετικής θέσης στις δύο εκδόσεις και ελέγχθηκε κατά πόσο διαφέρουν στις δύο λίστες. Συνολικά παρατηρήθηκε ότι οι δύο λίστες στις 1000 πιο συχνές λέξεις περιλαμβάνουν 895 κοινές λέξεις (ποσοστό όμοιων λέξεων 89,5%).

Για να κριθεί αν η παρατηρούμενη διαφοροποίηση είναι στατιστικά σημαντική επιλέχθηκε η ανάλυση τους με το μη παραμετρικό στατιστικό τεστ Wilcoxon Signed Rank Test (βλ. 'Wilcoxon Signed-Rank Test', 2011). Το συγκεκριμένο τεστ προτιμάται γιατί δεν προϋποθέτει συγκεκριμένη κατανομή στις ελεγχόμενες μεταβλητές. Το τεστ παίρνει υπόψη του το μέγεθος των διαφορών μεταξύ ζευγών τιμών και δίνει μεγαλύτερο βάρος σε ζεύγη που έχουν μεγάλες διαφορές από τα ζεύγη που έχουν μικρές διαφορές. Το τεστ επιβεβαίωσε ότι, τόσο στην σχετική θέση των 1000 συχνότερων λέξεων (z= -0,808, p= 0,41), όσο και στην σχετική συχνότητα εμφάνισής τους (z= -1,721, p= 0,08) δεν υπάρχει στατιστικά σημαντική διαφοροποίηση. Επομένως, ο διπλασιασμός του μεγέθους του ΕΘΕΓ δεν έχει επιδράσει ουσιαστικά, ούτε στην σχετική συχνότητα, ούτε στην σχετική θέση των 1000 συχνότερων λέξεων της ΝΕ.

Για να επιβεβαιωθεί το συγκεκριμένο αποτέλεσμα ελέγχθηκε η συνάφεια των δύο λιστών, τόσο ως προς την σχετική θέση, όσο και ως προς την σχετική συχνότητα των 1000 συχνότερων λέξεων. Ο συντελεστής συσχέτισης που χρησιμοποιήθηκε ήταν ο Spearman r (r_s). Η συνάρτηση του συντελεστή αυτού είναι:

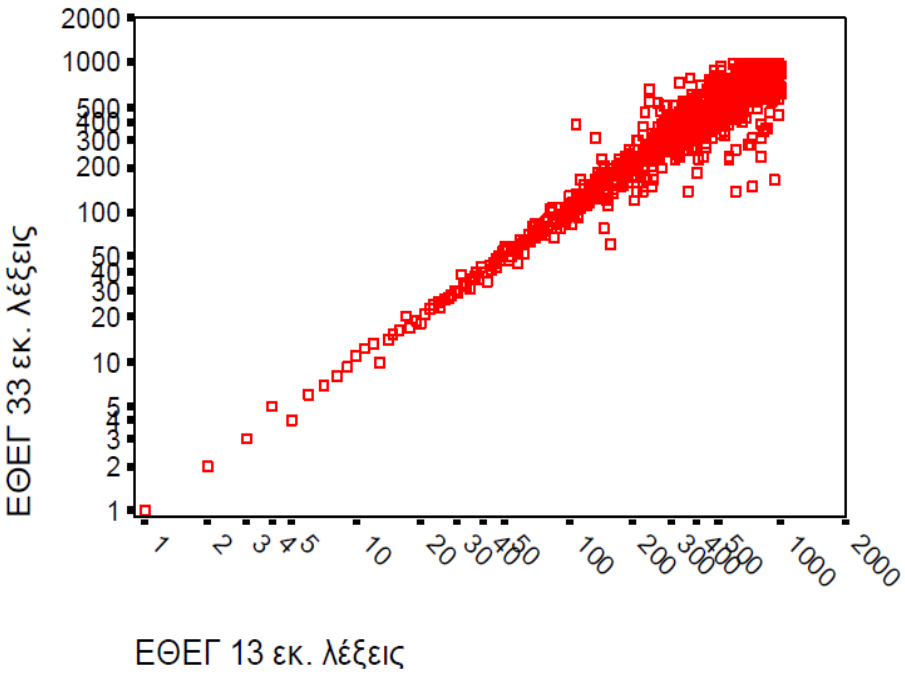
$$Rs=1-\frac{6Sigma*d^2}{n^3-n} \qquad (1.1)$$

Περισσότερες πληροφορίες για τον συντελεστή συσχέτισης Spearman βλ. ('Spearman's Rank Correlation', 2011).

Η συσχέτιση που παρατηρήθηκε ήταν υψηλότατη, τόσο για την σχέση της σχετικής συχνότητας των δύο εκδόσεων του ΕΘΕΓ (r_s= 0,89, p<0,001), όσο και για την σχέση της σχετικής θέσης (r_s= 0,90, p<0,001).

Στο διάγραμμα διασποράς (Διάγραμμα 1.2) (Μικρός υπό δημοσίευση) της σχετικής θέσης των 1000 συχνότερων λέξεων στις δύο εκδόσεις του ΕΘΕΓ δείχνει την

ισχυρή σταθερότητα που παρουσιάζει η σχετική θέση στο φάσμα των 1000 συχνότερων λέξεων της ΝΕ.



Διάγραμμα 2.2: Σύγκριση της σχετικής θέσης των 1000 συχνότερων λέξεων στις δύο εκδόσεις του ΕΘΕΓ.

Για να ολοκληρωθεί η σύγκριση των δύο εκδόσεων του ΕΘΕΓ πέραν των λέξεων αναλύθηκαν και οι λίστες των 1000 συχνότερων λημμάτων. Η σύγκριση έδειξε ότι η διαφοροποίηση είναι πολύ μεγαλύτερη από τις λίστες των λέξεων. Τα κοινά λήμματα στις δύο λίστες είναι 775 (ποσοστό όμοιων λημμάτων 77,5%). Η ανάλυση με Wilcoxon Signed Rank Test έδειξε μια στατιστικά σημαντική διαφοροποίηση, τόσο στις κατατάξεις των λημμάτων μέσα στις λίστες ($Z= -4,64$, $p<0,001$), όσο και στη σχετική συχνότητα εμφάνισης των λημμάτων ($Z= -2,16$, $p<0,05$).

1.5.2. Ο νόμος του Zipf στις λέξεις και τα λήμματα της ΝΕ.

Ένας από τους γνωστότερους νόμους στην ποσοτική γλωσσολογία είναι ο νόμος του Zipf (Rousseau & Zhang, 1992) που συνδέει την σχετική θέση μιας λέξης με την συχνότητά της. Αν και παρατηρήσεις για αυτήν την σχέση είχαν ήδη διαπιστωθεί στα τέλη του 19^{ου} αιώνα από τον Γάλλο ψυχολόγο Estoup (Těšitelová, 1992), ήταν ο Zipf που έκανε ευρύτερα γνωστή τη συγκεκριμένη παρατήρηση ως πρώτος νόμος του Zipf. Σύμφωνα με αυτόν η σχέση μεταξύ της σχετικής θέσης

μιας λέξης σε μια λίστα συχνότητας λέξεων (rank order) και της συχνότητας της είναι αντιστρόφως ανάλογη και το γινόμενό τους είναι σταθερό.

- Είναι ένας πολύ απλός εκθετικός νόμος (power law). Σχετίζει την σχετική θέση (r) ενός μέλους μιας διατεταγμένης λίστας με την συχνότητα εμφάνισης (p_i) αυτού του μέλους.:

$$P(i)= \frac{b}{i^a} \rightarrow \log(p(i))= B-a\log(i), \text{με } \alpha\approx 1 \qquad (1.2)$$

- Η εγκυρότητά του έχει παρατηρηθεί σε ένα μεγάλο εύρος φαινομένων, συμπεριλαμβανομένων των φυσικών γλωσσών, οικονομικών, οικολογικών συστημάτων και στατιστικών πρόσβασης σε δικτυακούς τόπους.

Ο πρώτος νόμος του Zipf έχει αποδειχθεί ότι ισχύει για πολλές γλώσσες (Miller, Newman, & Friedman, 1958);(Rousseau & Zhang, 1992). Η ερμηνεία αυτού του εμπειρικού νόμου έγκειται στην προσπάθεια του γλωσσικού συστήματος να εξισορροπήσει τη συχνότητας μιας λέξης με τον αριθμό των λέξεων που μοιράζονται την ίδια συχνότητα εμφάνισης. Η εξισορροπητική αυτή τάση σε κάθε γλωσσικό σύστημα είναι προϊόν ανταγωνισμού δύο αντίρροπων δυνάμεων. Η πρώτη ωθεί το γλωσσικό σύστημα να περιορίσει την λεξιλογική ποικιλία που σε θεωρητικό επίπεδο θα μπορούσε να φτάσει στη μία λέξη με τη μέγιστη δυνατή συχνότητα. Η αντίθετη δύναμη ωθεί το γλωσσικό σύστημα σε αύξηση της λεξιλογικής ποικιλότητας τείνοντας τη θεωρητική συχνότητα για κάθε λέξη στο 1.
Οι δύο αυτές δυνάμεις αντιστοιχούν σε αντίστοιχες απαιτήσεις που παρουσιάζονται στους εμπλεκόμενους σε κάθε επικοινωνιακό γεγονός. Ο πομπός θέλει να κωδικοποιεί το μήνυμα με την μικρότερη δυνατή προσπάθεια χρησιμοποιώντας τις ελάχιστες δυνατές λέξεις οι οποίες θα έχουν υψηλότατη συχνότητα. Αντίθετα, ο δέκτης θέλει τη μεγαλύτερη δυνατή πληροφορία από το μήνυμα που λαμβάνει, έτσι ώστε να απαιτείται η ελάχιστη δυνατή προσπάθεια αποκωδικοποίησης του.

1.6. ΤΟ ΜΗΚΟΣ ΛΕΞΗΣ ΣΤΗ ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ.

1.6.1. Η μελέτη των κατανομών του μήκους των λέξεων.

Το μήκος των λέξεων και η κατανομή τους έχουν αποτελέσει αντικείμενο εντατικής μελέτης στον χώρο της ποσοτικής γλωσσολογίας με σημαντικότερη πρωτοβουλία το πρόγραμμα Göttingen (Best, 1998). Η κατανομή του μήκους των λέξεων έχει εξεταστεί συγκριτικά για σχεδόν όλες τις ινδοευρωπαϊκές γλώσσες ήδη από τα μέσα του 20ου αιώνα με πρωτοπόρο τον Ρώσο μαθηματικό Čebanov (Altmann, 1988).

Σε μια από τις πρώτες συγκριτικές μελέτες για να βρεθεί αν το μήκος των λέξεων ακολουθεί συγκεκριμένη κατανομή ο Fucks (Fucks, 1956) εξετάζοντας δεδομένα από 8 Ινδοευρωπαϊκές και μη γλώσσες κατέληξε στην κατανομή «1 Displaced Poisson». Νεότερη έρευνα (Grotjahn, 1982) έδειξε ότι καταλληλότερη κατανομή για το μήκος των λέξεων είναι η «Negative Binomial», αφού δεν θεωρεί ότι οι πιθανότητες των μεμονωμένων λέξεων είναι ίσες, αλλά αναγνωρίζει την εξάρτησή τους από υφολογικούς, συμφραστικούς και άλλους παράγοντες.

Προσπάθεια για την μοντελοποίηση του μήκους των λέξεων έχει γίνει και από τον Altmann (Altmann, 1988) ο οποίος προσπαθεί να διασυνδέσει τις μαθηματικές κατανομές με τις αντίρροπες τάσεις που παρουσιάζει η γλώσσα ως επικοινωνιακό γεγονός όπως τις συνέλαβε ο Zipf στον πρώτο του νόμο. Η συγκριτική ανάλυση 38 γλωσσών που εκτείνονται σχεδόν στο σύνολο των γλωσσικών οικογενειών έδειξε ότι η κατανομή «Hyper-Poisson» είναι κατάλληλη για την περιγραφή του μήκους των λέξεων των περισσότερων γλωσσών (συμπεριλαμβανομένων και των αρχαίων ελληνικών) (Best, 1998).

Στη ΝΕ δεν έχει γίνει συστηματική έρευνα σχετικά με τα μήκη των λέξεων και τις κατανομές τους. Αν και το μέσο μήκος λέξεων σε ένα κείμενο αποτελεί σημαντικό υφομετρικό δείκτη ο οποίος έχει ήδη χρησιμοποιηθεί σε υφομετρικές αναλύσεις στη ΝΕ (Μπεκιάρη, Παπαβασιλείου, Πασχάλης 2001), καθώς και σε πειράματα κειμενικής κατηγοριοποίησης (Mikros & Carayannis, 2000), δεν έχει μελετηθεί διεξοδικά η κατανομή του. Στη συνέχεια θα παρατεθούν έρευνες για το μήκος των λέξεων του ΕΘΕΓ στο σύνολό του καθώς και στις 1000 συχνότερες λέξεις. Επιπλέον θα παρουσιαστούν ενδεικτικά η κατανομή του μήκους των

λέξεων σε ορισμένα κείμενα της ΝΕ και θα συγκρίνουμε τα κείμενα της ΝΕ με αντίστοιχες έρευνες σε άλλες γλώσσες.

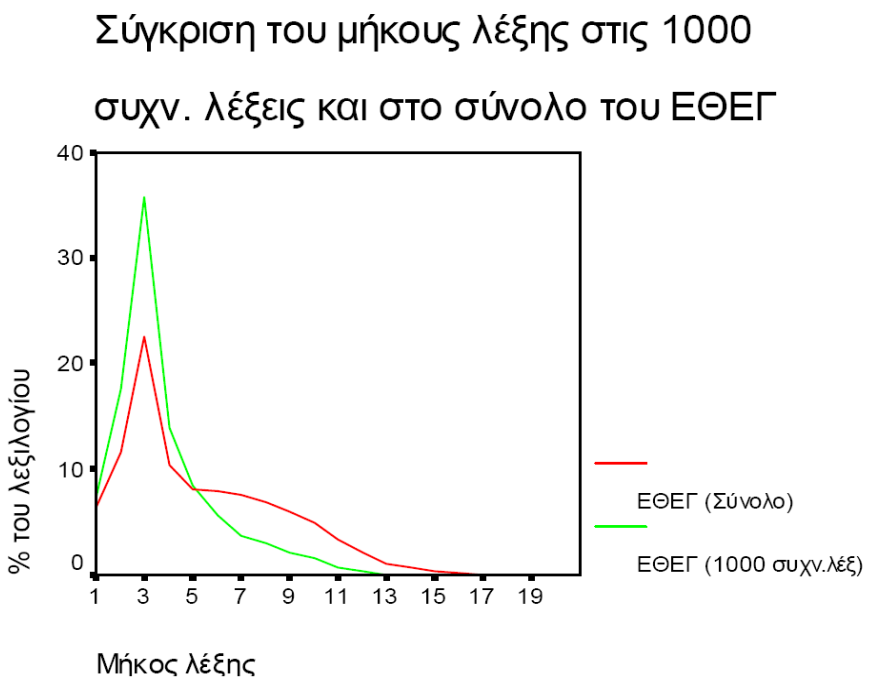
1.6.2. Μήκος λέξης στον ΕΘΕΓ.

Το μέσο μήκος της λέξης στο σύνολο του ΕΘΕΓ είναι 5,32 γράμματα. Ωστόσο, αυτός ο μέσος όρος δεν είναι ομοιογενής. Το μήκος των λέξεων αποτελεί μια ποσότητα που εξαρτάται από πολλούς παράγοντες ένας από τους οποίους είναι και το κειμενικό μέσο (Wimmer, et. al., 1994). Η επίδραση του συγκεκριμένου παράγοντα φαίνεται στον πίνακα 1.8 παρακάτω.

Πίνακας 2.7: Μέσο μήκος λέξης ανά κειμενικό μέσο στον ΕΘΕΓ.

ΕΘΕΓ (σύνολο)	Βιβλία	Εφημερίδες	Περιοδικά	Αδιευκρίνιστο
5,33	5,41	5,29	5,38	5,56

Διάγραμμα 2.3: Το μήκος των λέξεων στον ΕΘΕΓ στα διαφορετικά κειμενικά μέσα.



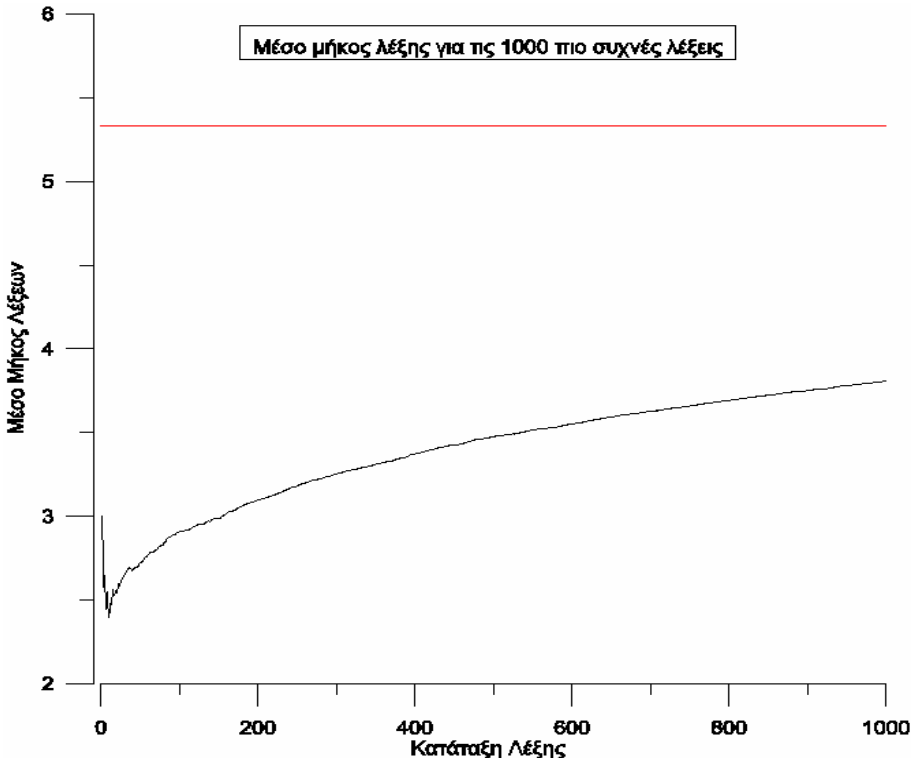
Διάγραμμα 2.4: Συγκριτικό διάγραμμα του μήκους των λέξεων στο σύνολο του ΕΘΕΓ και τις 1000 συχνότερες λέξεις του.

Η κατανομή των 1000 συχνότερων λέξεων επιδεικνύει μια συγκέντρωση στις μικρότερες λέξεις (1-5 γράμματα), ενώ στις μεσαίες και έπειτα φαίνεται ότι η κατανομή του συνόλου του ΕΘΕΓ καλύπτει μεγαλύτερη επιφάνεια. Επίσης ενδιαφέρον παρουσιάζει η μεταβολή του το αθροιστικού μέσου μήκους λέξεως για τις πιο συχνές λέξεις. Αν a_i είναι το μήκος της λέξης i και b_i είναι το σύνολο των εμφανίσεων της λέξης i τότε το αθροιστικό μέσο μήκος λέξεως είναι:

$$yi=\frac{\Sigma ai * bi}{\Sigma bi} \qquad (1.3)$$

Όπως παρατηρείται στο διάγραμμα 1.5, η ποσότητα αυτή αυξάνει μονοτονικά προσεγγίζοντας το μέσο μήκος λέξεων του ΕΘΕΓ που είναι 5,33 (ευθεία γραμμή). **Επιπλέον είναι εμφανές ότι οι συχνές λέξεις έχουν μικρότερο μήκος από το μέσο μήκος λέξεων του συνόλου του σώματος κειμένων**. Και οι δύο αυτές παρατηρήσεις συμφωνούν με αυτά που έχουν υπολογιστεί για άλλες γλώσσες (Grotjahn & Altmann 1993) και είναι συμβατά με την αρχή της «ελάχιστης προσπάθειας» του Zipf και την ευρύτερη αυτορυθμιστική ικανότητα των γλωσσικών συστημάτων που την εμφανίζουν σε όλα τα επίπεδα της οργάνωσής τους.

Τέλος, τα παραπάνω συμπεράσματα για το μήκος λέξεων είναι τα ίδια με αυτά που παρατηρήθηκαν με παλαιότερα στον ΕΘΕΓ των 13 εκ. λέξεων, αφού η μορφή της καμπύλης είχε την ίδια ακριβώς μορφή. Η μόνη αλλαγή είναι η μεταβολή του μέσου μήκους λέξεων από 5,45 που έχει υπολογιστεί παλαιότερα, σε 5,33 που υπολογίστηκε τώρα. Αυτή η μικρή μείωση δικαιολογείται από το γεγονός ότι άλλαξε η κατανομή της προέλευσης των κειμένων, με τις εφημερίδες, οι οποίες εμφανίζουν συστηματικά μικρότερο μήκος λέξεων (βλ. Πίνακας 1.8 παραπάνω), να καταλαμβάνουν πλέον ένα μεγαλύτερο ποσοστό απ' ότι στον ΕΘΕΓ 13 εκ.



Διάγραμμα 2.5: Αθροιστική αύξηση του μέσου μήκους λέξεων στις 1000 συχνότερες λέξεις.

Συμπληρωματικά με την μελέτη του μήκους των λέξεων σε μακροεπίπεδο, έχει ελεγχθεί (Mikros, Hatzigeorgiu, & Carayannis, 2005) η κατανομή του μήκους των λέξεων σε επίπεδο κειμένου. Για την σύγκριση της κατανομής του μήκους των λέξεων κειμένων επιλέχτηκαν τυχαία πέντε βιβλία ποικίλης θεματολογίας και αναλύθηκε η κατανομή του μήκους των λέξεων που αυτά παρουσίασαν. Η εξέταση των κατανομών έδειξε ότι τα δεδομένα μήκους λέξεων στη ΝΕ μπορούν να μοντελοποιηθούν ικανοποιητικά με την κατανομή Negative Binomial.

Η κατανομή Negative Binomial δίνεται από τον τύπο:

$$f(x)=\binom{s+x-1}{x} p^s (1-p)^x \qquad (1.4)$$

όπου

s=ο αριθμός των επιτυχιών, s>0

p=η πιθανότητα μιας επιτυχίας, 0<p<1

Τα αποτελέσματα της προσαρμογής φαίνονται στον πίνακα 1.9.

Πίνακας 2.8: Αποτελέσματα κατανομής Negative Binomial στην κατανομή μήκους των λέξεων σε 5 τυχαία κείμενα από τον ΕΘΕΓ.

	Παράμετροι κατανομής		Προσαρμογή	
Κείμενα	**s**	**P**	**X^2**	**$P(X^2)$**
Επιστημονική Μελέτη 1	**1**	**2,3E -4**	**4,41**	**0,21**
Επιστημονική Μελέτη 2	**1**	**5,6E -4**	**5,16**	**0,16**
Επιστημονική Μελέτη 3	**1**	**2,4E -4**	**3**	**0,39**
Διήγημα 1	**1**	**6,6E -4**	**4,83**	**0,18**
Διήγημα 2	**1**	**1,04E -4**	**4,82**	**0,18**
Νομολογία	**1**	**1,9E -4**	**5,16**	**0,15**

Η συγκεκριμένη κατανομή έχει χρησιμοποιηθεί και για την ερμηνεία των δεδομένων άλλων γλωσσών (Best, 1998; Wimmer & Altmann, 1996), αν και πολλές ευρωπαϊκές γλώσσες φαίνεται να ακολουθούν την Hyper-Poisson (Best, 1998). Τα συγκεκριμένα αποτελέσματα αποτελούν αφετηρία για μια πληρέστερη και μεθοδικότερη διερεύνηση των κατανομών που εφαρμόζουν στα ΝΕ κείμενα που θα ολοκληρωθεί στο μέλλον.

1.7. ΣΥΧΝΟΤΗΤΕΣ ΓΡΑΜΜΑΤΩΝ

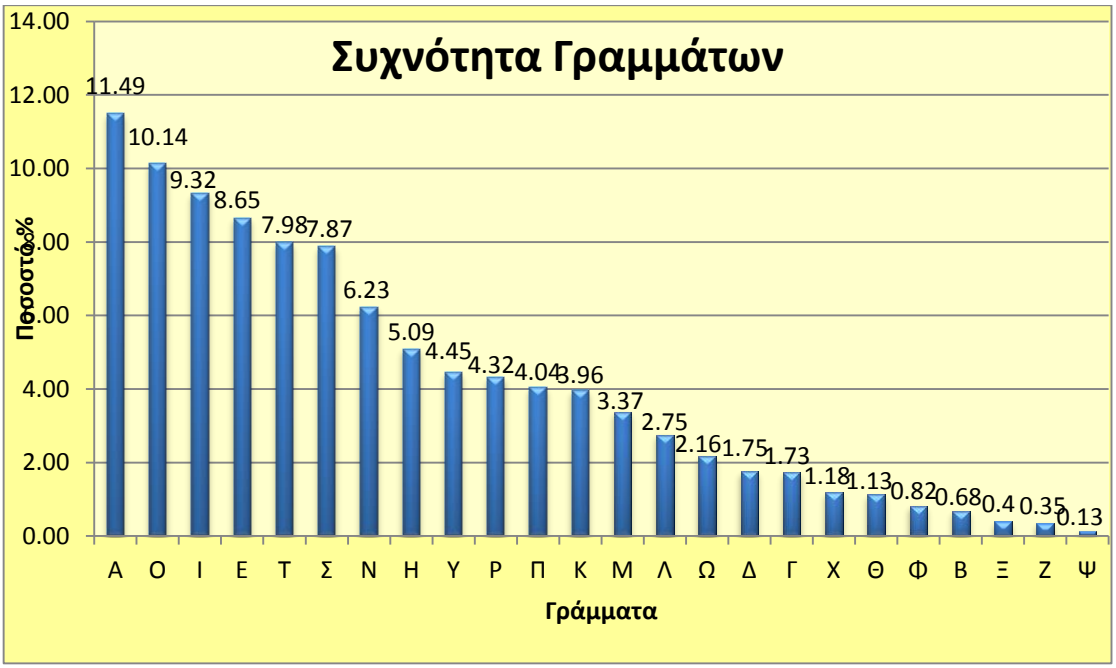
Οι ελληνικοί χαρακτήρες μετρήθηκαν σε όλο τον ΕΘΕΓ και το σύνολό τους είναι 151.235.762. Η κατανομή τους ανά συχνότητα εμφάνισης είναι η ακόλουθη που παρουσιάζεται στον πίνακα 1.10.

Πίνακας 2.10: Κατανομή της συχνότητας των γραμμάτων στον ΕΘΕΓ.

Γράμμα	Εμφανίσεις	Ποσοστό	Γράμμα	Εμφανίσεις	Ποσοστό
Α	18.990.738	**11,49**	**Μ**	5.571.176	**3,37**
Ο	16.756.541	**10,14**	**Λ**	4.549.986	**2,75**
Ι	15.399.917	**9,32**	**Ω**	3.576.349	**2,16**
Ε	14.287.940	**8,65**	**Δ**	2.891.994	**1,75**
Τ	13.182.878	**7,98**	**Γ**	2.859.998	**1,73**
Σ	13.010.111	**7,87**	**Χ**	1.953.016	**1,18**
Ν	10.297.844	**6,23**	**Θ**	1.860.300	**1,13**
Η	8.404.229	**5,09**	**Φ**	1.350.961	**0,82**
Υ	7.357.970	**4,45**	**Β**	1.124.308	**0,68**
Ρ	7.140.530	**4,32**	**Ξ**	668.976	**0,40**
Π	6.672.496	**4,04**	**Ζ**	573.490	**0,35**
Κ	6.540.793	**3,96**	**Ψ**	220.964	**0,13**

Η συγκεκριμένη κατανομή περιλαμβάνει τα τονισμένα, τα άτονα, τα κεφαλαία και τα πεζά γράμματα.

Στο διάγραμμα 1.6 απεικονίζεται η συχνότητα των γραμμάτων της ΝΕ σύμφωνα με την εμφάνισή τους στον ΕΘΕΓ.



Διάγραμμα 2.6: Κατάταξη των γραμμάτων της ΝΕ σύμφωνα με τη συχνότητα εμφάνισης τους στον ΕΘΕΓ.

Επιπλέον, εξετάστηκε η κατανομή των τονισμένων και των άτονων φωνηέντων η οποία φαίνεται στον πίνακα 1.11.

Πίνακας 2.9: Κατανομή τονισμένων και άτονων φωνηέντων στον ΕΘΕΓ.

Τονισμένα	Συχνότητα	%	Άτονα	Συχνότητα	%	Σύνολο
ά	3.310.673	17,43	α	15.680.065	82,57	18.990.738
έ	3.064.543	21,45	ε	11.223.397	78,55	14.287.940
ό	3.503.641	20,91	ο	13.252.900	79,09	16.756.541
ώ	1.204.832	33,69	ω	2.371.517	66,31	3.576.349
ί, ϊ, ΐ	4.033.474	26,17	ι	11.379.680	73,83	15.413.154
ή	2.207.937	26,27	η	6.196.292	73,73	8.404.229
ύ	1.658.234	22,54	υ	5.699.914	77,46	7.358.148

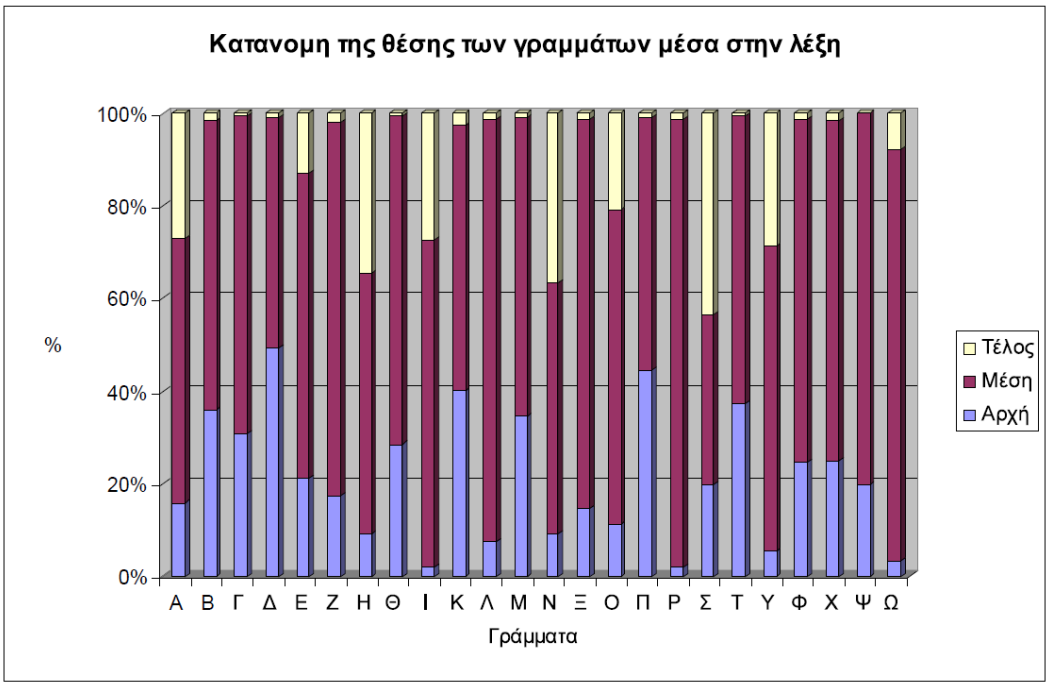
Τα τονισμένα φωνήεντα στο σύνολό τους αποτελούν σχεδόν το 1/3 των φωνηέντων και η αναλογία τους ποικίλει από το 1/2 (για το Ω) έως το 1/5 (για το Α).

Ακόμη εξετάστηκε η συχνότητα των ελληνικών γραμμάτων ανάλογα με την θέση τους στην λέξη. Τα αποτελέσματα δίνονται στον πίνακα 1.12.

Πίνακας 2.10: Συχνότητα των γραμμάτων ανάλογα με την θέση τους στην λέξη.

	Αρχή	%	Μέση	%	Τέλος	%	Σύνολο
Α	3.027.599	15,92	10.814.960	56,87	5.174.014	27,21	19.016.573
Β	406.797	35,79	710.905	62,55	18.863	1,66	1.136.565
Γ	889.836	30,92	1.965.217	68,28	22.950	0,80	2.878.003
Δ	1.434.827	49,22	1.451.962	49,81	28.439	0,98	2.915.228
Ε	3.055.110	21,35	9.385.872	65,60	1.867.108	13,05	14.308.090
Ζ	99.660	17,36	462.181	80,49	12.336	2,15	574.177
Η	826.501	9,19	5.051.230	56,14	3.119.571	34,67	8.997.302
Θ	533.658	28,59	1.318.805	70,64	14.404	0,77	1.866.867
Ι	291.969	1,89	10.924.053	70,85	4.201.458	27,25	15.417.480
Κ	2.648.806	40,00	3.803.065	57,43	170.615	2,58	6.622.486
Λ	336.735	7,40	4.148.222	91,10	68.385	1,50	4.553.342
Μ	1.940.285	34,67	3.605.974	64,43	50.468	0,90	5.596.727
Ν	960.140	9,30	5.578.176	54,00	3.791.231	36,70	10.329.547
Ξ	97.548	14,58	562.470	84,06	9.137	1,37	669.155
Ο	1.934.754	11,24	11.704.050	67,98	3.577.554	20,78	17.216.358
Π	2.970.971	44,42	3.650.980	54,58	67.159	1,00	6.689.110
Ρ	141.042	1,97	6.909.425	96,74	91.474	1,28	7.141.941
Σ	2.580.472	19,78	4.774.831	36,60	5.692.232	43,63	13.047.535
Τ	4.921.137	37,30	8.173.431	61,95	99.791	0,76	13.194.359
Υ	407.969	5,54	4.832.634	65,66	2.118.932	28,79	7.359.535
Φ	334.348	24,71	1.000.590	73,96	17.946	1,33	1,352.884
Χ	490.047	24,97	1.437.958	73,28	34.407	1,75	1.962.412
Ψ	44.054	19,93	176.695	79,92	333	0,15	221.082
Ω	119.559	3,34	3.177.579	88,82	280.330	7,84	3.577.486

Η συγκεκριμένη κατανομή αποτυπώνεται γραφηματικά στο διάγραμμα 1.7.



Διάγραμμα 2.7: Συγκριτικό διάγραμμα της κατανομής γραμμάτων μέσα στη λέξη.

Συνοψίζοντας, από τα παραπάνω προκύπτει ότι:

- ✓ **τα γράμματα Δ, Π, Κ και το Τ αποτελούν τα πιο συχνά σύμφωνα με τα οποία ξεκινούν ελληνικές λέξεις, ενώ τα Ε και Α είναι τα πιο συχνά αρχικά φωνήεντα**.
- ✓ **Αντίστοιχα οι πιο συχνοί χαρακτήρες στο τέλος της λέξης είναι το Σ και το Ν για τα σύμφωνα και το Η και του Υ για τα φωνήεντα.**
- ✓ **Τέλος τα σύμφωνα που εμφανίζονται σχεδόν κατ' αποκλειστικότητα στο εσωτερικό της λέξης είναι το Ρ και Λ με 97% και 91% αντίστοιχα.**

Σε μια πρόσφατη συγκριτική έρευνα της συχνότητας των χαρακτήρων σε πολλές ευρωπαϊκές γλώσσες οι Rosenbaum & Fleischmann (Rosenbaum & Fleischmann, 2002) υποστήριξαν ότι τα ΝΕ έχουν αυξημένη συχνότητα του κενού σε σχέση με τις ρωμανικές γλώσσες. **Ειδικότερα, υπολογίζεται το κενό στη ΝΕ σε 19,4% του συνόλου των χαρακτήρων του ΗΣΚ που εξετάστηκε, ενώ στη Λατινική και στις Ρωμανικές γλώσσες στο 14,6%. Ωστόσο, τα δεδομένα από τον ΕΘΕΓ δείχνουν ότι η συχνότητα του κενού στη Νέα Ελληνική γλώσσα ταυτίζεται πλήρως με τις ρωμανικές γλώσσες και είναι 14,6% επί του συνόλου των χαρακτήρων του ΕΘΕΓ (Mikros κ.ά., 2005).**

1.8. ΑΥΤΟΜΑΤΗ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΚΕΙΜΕΝΩΝ

Η αυτόματη κατηγοριοποίηση κειμένων (ΑΚΚ) αποτελεί μια από τις σημαντικότερες τεχνολογίες στην Επεξεργασία Φυσικής Γλώσσας (ΕΦΓ) και συνίσταται στην αυτόματη κατάταξη ενός κειμένου σε μία προκαθορισμένη θεματική κατηγορία. Τα τελευταία χρόνια έχει σημειωθεί ραγδαία αύξηση στις έρευνες ΑΚΚ αξιοποιώντας μεθόδους τόσο από την πολυπαραγοντική στατιστική , όσο και από το πεδίο της τεχνητής νοημοσύνης. Οι μέχρι τώρα προσπάθειες ΑΚΚ στη ΝΕ έχουν χρησιμοποιήσει μια ποικιλία πολυπαραγοντικών στατιστικών μεθόδων όπως Διακριτική Ανάλυση (Discriminant Function Analysis) (Mikros & Carayannis, 2000), Ανάλυση Συστάδων (Cluster Analysis) (Tambouratzis, Markantonatou, Hairetakis, & Carayannis, 2004) και Πολλαπλή Παλινδρόμηση (Multiple Regression) (Stamatatos et. al. 1999; Stamatatos 2000).
Επίσης, στις παραπάνω μελέτες έχει χρησιμοποιηθεί ένα ευρύ φάσμα μεταβλητών που εκτείνεται σε όλα τα γλωσσικά επίπεδα με αποτέλεσμα να μην είναι εφικτή η άμεση σύγκριση της αποτελεσματικότητας των χρησιμοποιούμενων στατιστικών μεθόδων και μεταβλητών που επιλέγονται. Για τους παραπάνω λόγους η χρήση ενός κοινού Ηλεκτρονικού Σώματος Κειμένων (ΗΣΚ) το οποίο θα αποτελέσει μια ενιαία βάση στην οποία θα συγκριθούν διαφορετικοί αλγόριθμοι και μεταβλητές.

1.8.1. Το Ηλεκτρονικό Σώμα Κειμένου εκπαίδευσης.

Σε μελέτη (Μικρός, 2005) που έγινε για την εξαγωγή στατιστικών συμπερασμάτων για τα ΗΣΚ στην εκπαίδευση, χρησιμοποιήθηκε ως βάση δεδομένων, δηλαδή ένα ΗΣΚ που αποτελείται από 900 άρθρα της «Ναυτεμπορικής» και χαρακτηρίζονται ως προς το θέμα που πραγματεύονται από την ίδια την εφημερίδα. Η κωδικοποίηση των θεμάτων και η κατανομή του αριθμού των λέξεων και του πλήθους των αρχείων δίνεται στον παρακάτω πίνακα 1.13.

Πίνακας 2.11: Ποσοτικά στοιχεία του ΗΣΚ εκπαίδευσης.

Κεντρικό θέμα	Αριθμός λέξεων	Μέσο μήκος άρθρου (λέξεις)	Τυπική Απόκλιση του μέσου μήκους	Μέγιστο μέγεθος	Ελάχιστο μέγεθος	Αριθμός στοιχείων
Κόσμος	33.692	225	136,6	81	969	150
Οικονομία	24.087	161	622,5	80	5748	150
Πολιτιστικά	26.976	180	106,8	83	739	150
Αθλητικά	35.136	234	124,4	81	1434	150
Πολιτικά	102.906	686	1499,7	81	5960	150
Media	31.395	209	117,5	80	671	150
Σύνολο	**254.192**	**282**	**692,9**			**900**

Το συγκεκριμένο ΗΣΚ παρουσιάζει ορισμένες ιδιαιτερότητες ως προς τη θεματική σύστασή του και επιλέχθηκε με τέτοιο τρόπο ώστε να δυσκολεύει την εργασία της ΑΚΚ. Τα βασικά σχεδιαστικά χαρακτηριστικά του που το διαφοροποιούν από τα προηγούμενα ΗΣΚ που χρησιμοποιήθηκαν για πειράματα ΑΚΚ είναι τα εξής:

1. Οι θεματικές κατηγορίες είναι εξαιρετικά «κοντινές» μεταξύ τους. Έτσι η κατηγορία «Media» είναι εξειδίκευση της κατηγορίας «Οικονομία», ενώ αντίστοιχη συνάφεια παρουσιάζουν θεματικά οι κατηγορίες «Κόσμος» και «Πολιτικά». Η θεματική συγγένεια των κειμένων ενός ΗΣΚ εκπαίδευσης αποτελεί ένα σημαντικό σημείο αξιολόγησης ενός αλγορίθμου κατηγοριοποίησης, αφού στις θεματικά «απομακρυσμένες» κατηγορίες οι περισσότεροι αλγόριθμοι επιδεικνύουν υψηλά ποσοστά ακρίβειας.
2. Η κατανομή των μεγεθών των άρθρων είναι εξαιρετικά ακανόνιστη γεγονός που αποκαλύπτεται και από την εξέταση των τυπικών αποκλίσεων ανά θεματική κατηγορία. Η μεγάλη πλειοψηφία των άρθρων (75%) είναι μικρού μεγέθους (<250 λέξεων), ενώ μόνο το 2% αποτελείται από άρθρα μεγαλύτερα των 1000 λέξεων. Αυτό έχει άμεσες επιπτώσεις στην σταθερότητα των μετρήσεων (Mikros 2002), αφού πολλά χαρακτηριστικά που επιλέγονται ως μεταβλητές δεν εμφανίζονται στο ΗΣΚ με αποτέλεσμα το αρχείο των δεδομένων να είναι σημαντικά ελλιπές. Επιπλέον, κειμενικά τμήματα μικρότερα των 500 λέξεων παρουσιάζουν έλλειψη σημαντικών

διακριτικών υφολογικών χαρακτηριστικών (Baillie,1974; Ledger & Merriam, 1994) και αυτό τα καθιστά δυσκολότερα κατηγοριοποιήσιμα έναντι μεγαλύτερων κειμένων. Το συγκεκριμένο χαρακτηριστικό αποτελεί πρόκληση για τους αλγόριθμους κατηγοριοποίησης οι οποίοι καλούνται να αντιμετωπίσουν πίνακες με ελλιπή πληροφορία γεγονός που προσομοιώνει τις συνθήκες λειτουργίας ενός συστήματος ΑΚΚ πραγματικού χρόνου.

3. Τα άρθρα προέρχονται όλα από την ίδια εφημερίδα και ανήκουν στο ίδιο υφολογικό επίπεδο λόγου (register) γεγονός που περιορίζει τις υφολογικές διαφοροποιήσεις αποκλειστικά στο θεματικό επίπεδο. Επιπλέον πιθανές κανονικοποιήσεις στο κείμενο προερχόμενες από τον εκδότη έχουν εφαρμοστεί ομοιογενώς και δεν στρεβλώνουν τις μετρήσεις (Rudman, 1997) σε αντίθεση με ένα ΗΣΚ που θα περιλάμβανε κείμενα από πολλές εφημερίδες όπου οι κανονικοποιήσεις είναι πολλές και διαφορετικού χαρακτήρα.

1.8.2. Μεταβλητές ενός ΑΚΚ.

Η επιλογή μεταβλητών για την ΑΚΚ αποτελεί από μόνη της σημαντικό ερευνητικό πεδίο με διαφορετικές προσεγγίσεις. Σε γενικές γραμμές οι μεθοδολογίες που έχουν χρησιμοποιηθεί εμπίπτουν στις παρακάτω γενικές κατηγορίες:

1. **Λεξιλογικές μέθοδοι:** Σε αυτή την κατηγορία μεθόδων χρησιμοποιούνται οι συχνότητες εμφάνισης κάποιων λέξεων ως ανεξάρτητες μεταβλητές οι οποίες αξιοποιούνται για την εκπαίδευση του αλγορίθμου κατηγοριοποίησης. Με την αυξημένη διαθεσιμότητα των ηλεκτρονικών γλωσσικών πόρων η ΑΚΚ έχει ενσωματώσει γλωσσική γνώση όπως αυτή έχει κωδικοποιηθεί σε Λεξικές Βάσεις Δεδομένων, όπως το WordNet (Buenaga, Gómez, & Díaz, 1997; Junker & Abecker, 1997; Scott & Matwin, 1999). Επίσης σε άλλες περιπτώσεις αξιοποιείται η ευρετική μεθοδολογία η οποία έχει αναπτυχθεί σε ορισμένα πεδία εφαρμογών της ΑΚΚ, όπως η αποφυγή ενοχλητικών ηλεκτρονικών μηνυμάτων (spam mail) (Gomez Hidalgo & de Buenaga Rodriguez, 1997; Sahami, Dumais, Heckerman, & Horvitz, 1998) . Μία άλλη ερευνητική προσέγγιση παίρνει υπ όψιν της την χρήση των συχνότερων λέξεων στο ΗΣΚ εκπαίδευσης. Η συγκεκριμένη τεχνική έχει αξιοποιηθεί με ιδιαιτέρα καλά αποτελέσματα στην αυτόματη

αναγνώριση συγγραφέα (Burrows, 1992; Burrows & Craig, 1994), αλλά έχει επεκταθεί και στην ΑΚΚ.

2. **Υφομετρικοί δείκτες:** Στα πλαίσια της υφομετρίας μια σειρά από μετρήσεις (μέσο μήκος λέξης, μέσο μήκος πρότασης κ.α.) σε κείμενα έχει αποδειχθεί ότι σχετίζονται με το προσωπικό ύφος των συγγραφέων. Οι συγκεκριμένοι δείκτες είναι γλωσσικά ανεξάρτητοι και αξιοποιούνται με επιτυχία στην αναγνώριση του γλωσσικού ύφους μεμονωμένων συγγραφέων. Στο μέτρο που το γλωσσικό ύφος καθορίζεται εν μέρει από το θέμα του κειμένου (Karlgren, 1999) οι συγκεκριμένοι δείκτες μπορούν να βελτιώσουν σημαντικά την απόδοση συστημάτων ΑΚΚ όταν συνδυαστούν με λεξικές μεταβλητές (Forsyth & Holmes, 1996).
3. **Γλωσσικά εξαρτημένες μεταβλητές:** Οι ιδιαιτερότητες της κάθε γλώσσας μπορούν να χρησιμοποιηθούν με τέτοιο τρόπο ώστε να αυξηθεί η απόδοση της ΑΚΚ. Έχει παρατηρηθεί π.χ. ότι γλώσσες με πλούσιο μορφολογικό σύστημα μπορούν να αξιοποιήσουν με ιδιαίτερη επιτυχία μικρότερες της λέξης μονάδες (Hoch, 1994).

1.9. ΕΡΕΥΝΑ ΣΤΟ SCRABBLE

Αναφέροντας στις παραπάνω παραγράφους την ανάλυση της γλώσσας σχετικά με την κατανομή των γραμμάτων και την συχνότητα εμφάνισής τους στον γραπτό λόγο, θεωρείται σκόπιμο να αναφερθεί ένα παράδειγμα της καθημερινής ζωής. Αυτό αφορά το παιχνίδι «scrabble», το οποίο έχει ο στόχο την κατασκευή και δημιουργία λέξεων. Στην εικόνα 1.1 απεικονίζεται το παιχνίδι «scrabble».

Εικόνα 2.1: Το παιχνίδι scrabble.

Αυτό το οποίο όμως πρέπει να αναλυθεί είναι ότι για την δημιουργία των λέξεων οι παίχτες κάνουν χρήση ενός συγκεκριμένου αριθμού γραμμάτων. Ο αριθμός αυτός κυμαίνεται στα 95 με 110 γράμματα ανάλογα με την γλώσσα των γραμμάτων και την έκδοση του παιχνιδιού. Μέσω αυτής της έρευνας θα γίνει αναφορά για την αγγλική και την ελληνική γλώσσα των οποίων η ανάλυση κατανομής γραμμάτων αναφέρθηκε στις παραπάνω παραγράφους. Πρέπει να τονιστεί ότι σε όλες τις γλώσσες η κατανομή των γραμμάτων δεν γίνεται με τυχαίο τρόπο αλλά βασίζονται σε λεξικά προκειμένου να είναι δυνατή η ροή του παιχνιδιού.

Το scrabble παρουσιάζει συγκεκριμένη τακτική στον τρόπο παιχνιδιού και υπάρχουν σαφείς κανόνες για την συχνότητα των γραμμάτων του (Appel & Jacobson, 1988). Στην παρακάτω εικόνα εμφανίζονται τα γράμματα του παιχνιδιού, στην αγγλική γλώσσα με τέτοιο τρόπο που προσδιορίζουν και το πλήθος των γραμμάτων.

Εικόνα 2.2: Τα γράμματα του scrabble.

Έτσι λοιπόν τα γράμματα στο scrabble εμφανίζονται με ίδιο ποσοστό εμφάνισης τους στο σύνολο των γραμμάτων που αποτελούν το παιχνίδι «scrabble», με τα ποσοστά εμφάνισης των αγγλικών γραμμάτων σε κείμενα και λεξικά (‘Oxford Dictionaries Online’, 2011). Δηλαδή το scrabble με αγγλικά γράμματα έχουν ποσοστό που δείχνει ο παρακάτω πίνακας 1.14.

Πίνακας 2.12: Συχνότητα γραμμάτων στο scrabble.

Γράμμα	Πλήθος	Ποσοστό(%)	Γράμμα	Πλήθος	Ποσοστό(%)
E	12	12,244	C	2	2,040
A	9	9,183	M	2	2,040
I	9	9,183	P	2	2,040
O	8	8,163	F	2	2,040
N	6	6,122	H	2	2,040
R	6	6,122	V	2	2,040
T	6	6,122	W	2	2,040
L	4	4,081	Y	2	2,040
S	4	4,081	K	1	1,020
U	4	4,081	J	1	1,020
D	4	4,081	X	1	1,020
G	3	3,061	Q	1	1,020
B	2	2,040	Z	1	1,020
Σύνολο Γραμμάτων: 98					

Για το scrabble με ελληνικά γράμματα τα δεδομένα αλλάζουν και παρουσιάζονται στον παρακάτω πίνακα 1.15.

Πίνακας 2.13: Ελληνικά γράμματα στο scrabble.

Γράμμα	Πλήθος	Ποσοστό(%)	Γράμμα	Πλήθος	Ποσοστό(%)
Α	12	11,764	Λ	3	2,941
Ο	9	8,823	Μ	3	2,941
Ε	8	7,843	Ω	3	2,941
Ι	8	7,843	Γ	2	1,960
Τ	8	7,843	Δ	2	1,960
Η	7	6,862	Β	1	0,980
Σ	7	6,862	Φ	1	0,980
Ν	6	5,882	Χ	1	0,980
Ρ	5	4,901	Ζ	1	0,980
Κ	4	3,921	Θ	1	0,980
Π	4	3,921	Ξ	1	0,980
Υ	4	3,921	Ψ	1	0,980

Έχοντας επίσημα αποτελέσματα από έρευνες που έχουν δημοσιευτεί και για την αγγλική και για την ελληνική γλώσσα μπορεί να γίνει μια σύγκριση των τιμών αυτών από την ανάλυση που έγινε από τις συχνότητες των γραμμάτων που υπολογίστηκαν στο παιχνίδι του scrabble.

Συγκεκριμένα, για την ελληνική γλώσσα, η οποία αποτελεί και το κύριο αντικείμενο της μελέτης, συμπεραίνεται ότι συγκρίνοντας τα ποσοστά ελληνικών γραμμάτων που εμφανίζονται στα ελληνικά κείμενα, όπως αυτά

έχουν ανακοινωθεί από τον ΕΘΕΓ ('ΕΘΕΓ', 2011) υπάρχει κατά προσέγγιση ταύτιση των ποσοστών με το ποσοστό γραμμάτων στο παιχνίδι scrabble. Δηλαδή όπως φαίνεται και από τον πίνακα 1.16 παρακάτω, τα πιο συχνά γράμματα παρουσιάζονται με την ίδια σειρά με ελάχιστες ποσοστιαίες διαφοροποιήσεις. Πρέπει όμως να ληφθεί υπ' όψιν ότι ο ΕΘΕΓ εξειδικεύει την έρευνά του σε μεγάλο όγκο δεδομένων ενώ στο scrabble τα δεδομένα είναι πολύ λιγότερα και συνεπώς μια τέτοια διαφοροποίηση είναι αναμενόμενη.

Πίνακας 2.14: Σύγκριση συχνότερων ελληνικών γραμμάτων στον ΕΘΕΓ και στο SCRABBLE.

α/α	Γράμμα	Ποσοστό (%) στον ΕΘΕΓ	Ποσοστό (%) στο Scrabble
1	Α	11,49	11,76
2	Ο	10,14	8,82
3	Ι	9,32	7,84
4	Η	5,09	6,86
5	Ε	8,65	7,84
6	Τ	7,98	7,84
7	Σ	7,87	6,86
8	Ν	6,23	5,88

1.10. ΣΥΜΠΕΡΑΣΜΑΤΑ

Από όλες τις παραπάνω παραγράφους συνοψίζεται ότι για την αγγλική γλώσσα υπάρχει πλούσιο υλικό που προσδιορίζει ξεκάθαρα την ανάλυση της γλώσσας. Για την ελληνική γλώσσα υπάρχουν έρευνες του ΕΘΕΓ, ο οποίος για πρώτη φορά στην Ελλάδα διεξήγαγε έρευνα σχετικά με την ανάλυση της ελληνικής γλώσσας. Με τη βοήθεια του ΕΘΕΓ προέκυψαν συμπεράσματα που έως τώρα δεν ήταν αρκετά σαφή. Παρουσιάστηκαν τα συχνότερα ελληνικά γράμματα και τον βαθμό στον οποίο επηρεάζεται η συχνότητά τους από την θέση τους στον γραπτό λόγο. Ιδιαίτερο ρόλο διαδραματίζει για ένα γράμμα αν αυτό τονίζεται, αν τοποθετείται στην αρχή ή στην μέση μιας λέξης, αλλά ακόμη περισσότερο επηρεάζονται και από το είδους του κειμένου που αυτά συμπεριλαμβάνονται.

1.11. ΕΠΙΛΟΓΟΣ

Μέσα από την εξέταση των ερευνών διαπιστώθηκαν κάποια βασικά στοιχεία της γραπτής ΝΕ γλώσσας και εξήχθησαν ορισμένες περιγραφικές παρατηρήσεις σχετικά με το συχνόχρηστο λεξιλόγιο (100 συχνότερες λέξεις), καθώς και τις συχνότερες των ελληνικών γραμμάτων. Οι παραπάνω μετρήσεις αποτυπώνουν μερικά από τα βασικότερα ποσοτικά χαρακτηριστικά της ΝΕ γραπτής γλώσσας και, αν και κάποιες από αυτές έχουν γίνει παλιότερα, είναι οι πρώτες που στηρίζονται σε ένα μεγάλο ΗΣΚ γενικής γλώσσας, τον ΕΘΕΓ.

Συνοψίζοντας, προκύπτει το συμπέρασμα ότι **στην ελληνική γλώσσα τα συχνότερα γράμματα σύμφωνα με τον ΕΘΕΓ είναι τα Α, Ι, Ο, Ε και Τ.**

Στο επόμενο κεφάλαιο θα μελετηθεί επιμέρους η ανάλυση της δομής της γλώσσας, ξεκινώντας από την γενική έννοια της κρυπτανάλυσης και αναλύοντας επιμέρους εφαρμογές της κρυπτογραφίας.

ΚΕΦΑΛΑΙΟ 2

2. ΚΡΥΠΤΟΓΡΑΦΙΑ ΚΑΙ ΚΡΥΠΤΑΝΑΛΥΣΗ

2.1. ΕΙΣΑΓΩΓΗ

Σε αυτό το κεφάλαιο εξετάζεται η **κρυπτολογία** (**cryptology)** η οποία χωρίζεται σε δύο μεγάλες κατηγορίες, την **κρυπτογραφία (cryptography)** και την **κρυπτανάλυση (cryptanalysis)**. Αρχικά αναλύονται και δίνονται ορισμοί οι οποίοι σχετίζονται με αυτές τις έννοιες και θα χρησιμοποιηθούν ευρέως σε αυτό το κεφάλαιο. Ιδιαίτερη έμφαση θα δοθεί συγκεκριμένα στην κατηγορία της κρυπτογραφίας η οποία θα αναλυθεί εκτενώς. Παρουσιάζεται η ιστορική αναδρομή της κρυπτογραφίας και θα παρουσιαστούν επίσης οι διάφορες κατηγορίες αλγορίθμων που ανήκουν σε αυτή.
Συγκεκριμένα αναλύονται και παρουσιάζονται οι αλγόριθμοι **Data Encryption Standard (DES**) και **Advanced Encryption Standard (AES)** οι οποίοι αποτελούν τους πιο διαδεδομένους ακόμη και σήμερα. Με τη βοήθεια αυτών των αλγορίθμων παρουσιάζεται και αναλύεται ο τρόπος της κρυπτογράφησης. Μέσω αυτών, αυτό το κεφάλαιο δίνει μια βαθύτερη ανάλυση σχετικά με τον τρόπο κατανομής των γραμμάτων αλλά και κατά πόσο αυτό επηρεάζει την συχνότητα και την εμφάνισή τους σε ένα κείμενο αλλά και έπειτα από την διαδικασία κρυπτογράφησης (encryption) και αποκρυπτογράφησης (decryption), εφαρμογή που θα παρουσιαστεί στη συνέχεια του κεφαλαίου.

2.1.1. Πρόλογος

Η λέξη **κρυπτολογία (cryptology)** αποτελείται από την ελληνική ρίζα κρύπτο-κρυφός και την λέξη λόγος. Είναι ο τομέας που ασχολείται με την μελέτη της ασφαλούς επικοινωνίας. Ο κύριος στόχος είναι να παρέχει μηχανισμούς για δύο ή και περισσότερα μέλη να επικοινωνήσουν χωρίς κάποιος άλλος να είναι ικανός να διαβάζει την πληροφορία εκτός από τα μέλη.
Η **κρυπτολογία** χωρίζεται σε δύο επιμέρους ενότητες (Ζορκάδης, 2002):

1. Την κρυπτογραφία

2. Την κρυπτανάλυση

Η **κρυπτογραφία** είναι ο επιστημονικός κλάδος που πραγματεύεται τη μελέτη και σχεδίαση κρυπτογραφικών τεχνικών, συστημάτων και πρωτοκόλλων (Ζορκάδης 2002). «Επίσης η **κρυπτογραφία** ορίζεται ως η μελέτη των μαθηματικών τεχνικών που σχετίζονται με πλευρές της ασφάλειας πληροφοριών όπως η εμπιστευτικότητα, η ακεραιότητα των δεδομένων, η πιστοποίηση αυθεντικότητα οντότητας και η πιστοποίηση αυθεντικότητας της πηγής δεδομένων» (Menezes, et. al., 1996). Ένας πιο πλήρης ορισμός δόθηκε από τον Rivest το 1990 όπου εισάγει την έννοια του αντίπαλου και ορίζει «η **κρυπτογραφία** ασχολείται με την επικοινωνία παρουσία αντιπάλου» (Τσιάκης 2005).

Μαζί με τον κλάδο της **κρυπτανάλυσης,** που ασχολείται με τη μελέτη τρόπων παραβίασης αυτών, απαρτίζουν την επιστήμη της Κρυπτολογίας. Έτσι, Κρυπτολογία είναι η επιστήμη της απόκρυψης, από τη μία πλευρά και από την άλλη πλευρά, της αποκάλυψης του περιεχομένου κωδικοποιημένων μηνυμάτων ή δεδομένων (Ζορκάδης 2002).

Η επιθυμία προστασίας του περιεχομένου μηνυμάτων οδήγησε στην επινόηση και χρήση κρυπτογραφικών τεχνικών και συστημάτων τα οποία επιτρέπουν το μετασχηματισμό μηνυμάτων ή δεδομένων κατά τέτοιο τρόπο ώστε να είναι αδύνατη η υποκλοπή του περιεχομένου τους κατά τη μετάδοση ή αποθήκευσή τους και, βεβαίως, την αντιστροφή του μετασχηματισμού. Η διαδικασία μετασχηματισμού καλείται **κρυπτογράφηση** και η αντίστροφή της, δηλαδή η επαναφορά του αρχικού κειμένου από το κρυπτογραφημένο κείμενο, **αποκρυπτογράφηση** (Ζορκάδης 2002; Γιαννακόπουλος 2007).

2.1.2. Ιστορική αναδρομή

Με βάση τις γενικές πληροφορίες και γνώσεις που μας παρέχονται σήμερα μπορεί να παρατεθεί μία σύντομη ιστορική αναδρομή σχετικά με την εξέλιξη της κρυπτογραφίας. «*Η κρυπτογραφία έχει μακρά και εντυπωσιακή ιστορία*». (Menezes, et.al., 1996). Οι ρίζες της ξεκινούν πριν από 4000 χρόνια περίπου, όταν στην αρχαία Αίγυπτο το 2000 π.Χ. χρησιμοποιήθηκαν τα ιερογλυφικά για να διακοσμήσουν τους τάφους εκλιπόντων κυβερνώντων ή βασιλέων. Τα ιερογλυφικά αυτά εξιστορούσαν την ζωή του εκλιπόντος και εξυμνούσαν τις σπουδαίες πράξεις του. Ήταν σκόπιμα περίπλοκη γραφή χωρίς όμως να σκοπεύει στην απόκρυψη

του κειμένου. Αντίθετα, σκοπός της ήταν να αποδώσει στο κείμενο σπουδαιότητα. Με το πέρασμα του χρόνου τα ιερογλυφικά γίνονταν όλο και περισσότερο πολύπλοκα ώσπου τελικά οι άνθρωποι έχασαν το ενδιαφέρον της αποκρυπτογράφησής τους. Στην Ινδία η μυστική γραφή εξελίχθηκε περισσότερο καθώς η κυβέρνηση χρησιμοποιούσε μυστικούς κώδικες για την επικοινωνία με ένα δίκτυο κατασκόπων διασκορπισμένων σε ολόκληρη τη χώρα. Οι κώδικες αυτοί αποτελούνταν κυρίως από απλές αλφαβητικές αντικαταστάσεις που βασίζονταν στα φωνήεντα. Η ιστορία της κρυπτογραφίας στην Μεσοποταμία είναι παρόμοια με αυτήν της Αιγύπτου καθώς η σφηνοειδής γραφή χρησιμοποιήθηκε για την κρυπτογράφηση κειμένου.

Η πιο ολοκληρωμένη μελέτη πάνω στο θέμα της κρυπτογραφίας έγινε από τον Kahn (Kahn 1974) ο οποίος αναλύει τη εξέλιξη της κρυπτογραφίας από την ελάχιστη και περιορισμένη χρήση της από τους Αιγύπτιους, 4000 χρόνια πριν και καταλήγει στη σύγχρονη εποχή. Το βιβλίο του κατάφερε να ολοκληρωθεί στο 1963, όπου και κάλυψε όλα τα σπουδαιότερα ζητήματα έως εκείνη την εποχή.

Στην Ελλάδα συγκεκριμένα, τα πρώτα στοιχεία κρυπτογραφίας εμφανίζονται στο κείμενο της Ιλιάδας, στην ιστορία του Ηροδότου αλλά και στην ιστορία του Περίανδρου της Κορίνθου (Λάσκαρη 2010). Η γνωστή «σκυτάλη» της Σπάρτης ανάγεται στον 5ο π.Χ. αιώνα. Αποτελούνταν από μία ξύλινη ράβδο γύρω από την οποία τύλιγαν μία λωρίδα περγαμηνής ή δέρματος (Εικόνα 2.1).

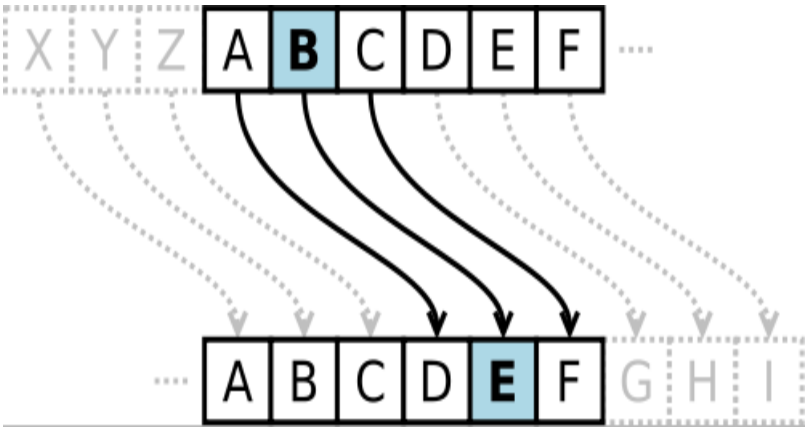
Εικόνα 2.1: Η Σπαρτιάτικη σκυτάλη, μια πρώιμη συσκευή για την κρυπτογράφηση.

Το μήνυμα γραφόταν από τον αποστολέα πάνω στη λωρίδα κατά μήκος της σκυτάλης και στη συνέχεια ξετύλιγαν τη λωρίδα η οποία φαινόταν να περιέχει μία σειρά γραμμάτων χωρίς νόημα καθώς το μήνυμα είχε αναδιαταχθεί. Προκειμένου να διαβάσει το μήνυμα ο παραλήπτης έπρεπε απλώς να τυλίξει την λωρίδα σε μια

σκυτάλη ίδιας διαμέτρου με αυτήν που χρησιμοποιήθηκε κατά την γραφή (Εικόνα 2.2).

Εικόνα 2.2: Η σκυτάλη στην οποία υπάρχει το κρυπτογραφημένο μήνυμα.

Με την κρυπτογραφική μέθοδο της σκυτάλης ο Λύσανδρος το 404 π.Χ. απέκρουσε με επιτυχία μία επίθεση του Πέρση σατράπη Φαρνάβαζου. Τον 4ο π.Χ. αιώνα ο Αινείας ο Τακτικός από την Μεγαλόπολη έγραψε ένα από τα πρώτα βιβλία κρυπτολογίας στο οποίο περιέχεται και μια τεχνική, παραλλαγή της οποίας εφαρμόστηκε από τους Γερμανούς στο Β΄ παγκόσμιο πόλεμο. Επίσης, ο Πολύβιος τον 2ο π.Χ. αιώνα δημιούργησε ένα κρυπτοσύστημα που ονομάστηκε το «τετράγωνο του Πολύβιου». Στο σύστημα αυτό κάθε γράμμα του κειμένου αντικαθίσταται από ζεύγη αριθμών με βάση έναν πίνακα διάστασης 5×5. Ο Ιούλιος Καίσαρας χρησιμοποίησε τη μυστική αυτή γραφή αλλά και κρυπτογράφηση μετατόπισης κατά τρεις χαρακτήρες δεξιά (Εικόνα 2.3).



Εικόνα 2.3: Ο κωδικοποιητής του Καίσαρα, μετατόπισης τριών θέσεων των γραμμάτων.

Κατά την περίοδο 400-1200 μ.Χ. η Ευρώπη διανύει τον μεσαίωνα κατά τον οποίο όλες οι επιστήμες και μεταξύ αυτών και η κρυπτογραφία ήταν σε παρακμή. Έτσι, την περίοδο αυτή η κρυπτογραφία αναπτύσσεται μόνο στην Ινδία και στις ισλαμικές χώρες. Τον 9ο μ.Χ. αιώνα ο πλέον σημαντικός Άραβας κρυπτολόγος Al

Kindi ανέπτυξε τη μέθοδο κρυπτανάλυσης που βασίζεται στη συχνότητα εμφάνισης των χαρακτήρων κάθε γλώσσας.

Το 1404 μ.Χ. ο Ιταλός αρχιτέκτονας και συγγραφέας L.B. Alberti ήταν ο πρώτος που ανέπτυξε ένα πολυαλφαβητικό σύστημα αντικατάστασης για κρυπτογράφηση. Για την υλοποίηση αυτού του κρυπτοσυστήματος επινόησε και την πρώτη κρυπτογραφική μηχανή μετά την σκυτάλη που ονομάζεται «δίσκοι του Alberti». Η μηχανή αυτή περιλάμβανε δύο ομόκεντρους χάλκινους δίσκους διαφορετικής διαμέτρου πάνω στην περιφέρεια των οποίων ήταν χαραγμένο ένα αλφάβητο. Ο ένας δίσκος αναφερόταν στο αρχικό μήνυμα και ο άλλος στο κρυπτογράφημα. Κάθε δίσκος περιστρεφόταν ανεξάρτητα και έτσι διαφορετικά μέρη του μηνύματος κρυπτογραφούνταν θέτοντας σε διαφορετικές θέσεις τον κρυπτογραφημένο δίσκο.

Το 1518 μ.Χ. ο Trithemius έγραψε την «Πολυγραφία», το πρώτο τυπωμένο βιβλίο κρυπτογραφίας, το οποίο περιείχε τα πρώτα κρυπτοσυστήματα κυκλικών μεταθέσεων και στη συνέχεια ο Cardano δημιούργησε την κρυπτογράφηση με τη μέθοδο μάσκας. Αυτή η κλάση κρυπτοσυστημάτων παρέμεινε σθεναρή μέχρι το 1800.

Τα μεταγενέστερα χρόνια η κρυπτογραφία άρχισε να εξελίσσεται. Όλες οι κυβερνήσεις της Δυτικής Ευρώπης χρησιμοποιούσαν την κρυπτογραφία με διάφορους τρόπους. Έως το 1860 μεγάλοι κώδικες ήταν πλέον συνήθεις για την κρυπτογράφηση των διπλωματικών επικοινωνιών. Η εφεύρεση του τηλέγραφου και του ραδιόφωνου εξώθησε την ανάπτυξη της κρυπτογραφικής προστασίας των τηλεπικοινωνιών, καθώς η ταχύτητα και το μέγεθος των μεταδιδόμενων δεδομένων κατέστησαν τις τηλεπικοινωνίες περισσότερο ευάλωτες στις μεσολαβήσεις και στην αποκρυπτογράφηση.

Τον 19ο αιώνα ο Ch. Babbage κατάφερε να κρυπταναλύσει διγραφικά κρυπτογραφήματα, δηλαδή συστήματα όπου κρυπτογραφούνται ζεύγη γραμμάτων. Το 1918 ο Γερμανός A. Serbious δημιούργησε ένα μηχανικό κρυπτοσύστημα που μπορεί να θεωρηθεί ως ηλεκτρική παραλλαγή των δίσκων του Alberti, και είναι γνωστή με το όνομα «Αίνιγμα» (Εικόνα 2.4).

Εικόνα 2.4: Η μηχανή Αίνιγμα χρησιμοποιήθηκε ευρέως στη Γερμανία.

Η βασική εκδοχή του Αινίγματος αποτελείται από ένα πληκτρολόγιο για την εισαγωγή του κειμένου, από μια αναδιατακτική μονάδα που κρυπτογραφεί κάθε χαρακτήρα του κειμένου σε έναν αντίστοιχο χαρακτήρα του κρυπτογραφήματος και έναν ηλεκτρικό πίνακα που δείχνει τον αντίστοιχο χαρακτήρα του τελικού κρυπτογραφημένου κειμένου. Στη βασική αυτή εκδοχή του κρυπτοσυστήματος σύντομα προστέθηκαν περισσότεροι αναδιατάκτες και ανακλαστές καθιστώντας την διαδικασία αρκετά πολύπλοκη μέχρι το Αίνιγμα να φτάσει να έχει πάνω από 10^{16} πιθανά κλειδιά. Ωστόσο, το 1938 ο Πολωνός μαθηματικός M. Rajewski με την ομάδα του κατάφεραν να σπάσουν το Αίνιγμα στην απλή μορφή που τότε είχε, και τα αποτελέσματά τους βοήθησαν μεταγενέστερα στην πλήρη κρυπτανάλυση του Αινίγματος από Άγγλους κρυπταναλυτές.

Τα μεταπολεμικά χρόνια η κρυπτογραφία γνώρισε τη μεγαλύτερη ανάπτυξη κυρίως λόγω της τεχνολογίας των υπολογιστών και της δυνατότητας ταχύτερων υπολογισμών. Ωστόσο, τα κλασικά κρυπτοσυστήματα που χρησιμοποιούνταν χαρακτηρίζονταν από το πρόβλημα της μυστικής ανταλλαγής κλειδιών, δηλαδή την δημιουργία ενός ασφαλούς διαύλου επικοινωνίας μέσω του οποίου ο αποστολέας και ο παραλήπτης θα μπορούσαν να συμφωνήσουν για το χρησιμοποιούμενο κλειδί. Εκτός από το πρόβλημα της ασφάλειας του διαύλου επικοινωνίας, υπήρχαν περιπτώσεις όπου η επικοινωνία έπρεπε να είναι γρήγορη, φθηνή και να βοηθά στην συχνή ανταλλαγή κλειδιών.

Το 1874 στο βιβλίο του W. S. Jevons (Jevons 1958) περιγράφηκε για πρώτη φορά η σχέση των μονόδρομων συναρτήσεων με την κρυπτογραφία., που αποτέλεσε τον πρόδρομο του δημόσιου κλειδιού.

Το 1976 οι W. Diffie και M. Hellman πρότειναν μια καινοτόμο μέθοδο κρυπτογράφηση σύμφωνα με την οποία μπορούσε να υπάρξει επικοινωνία χωρίς προηγούμενη ανταλλαγή κλειδιών μέσω ασφαλούς διαύλου. Η ασφάλεια της μεθόδου ανταλλαγής κλειδιών που πρότειναν βασιζόταν στο πρόβλημα επίλυσης του διακριτού λογαρίθμου και έγινε γνωστή ως ανταλλαγή κλειδιών Diffie-Hellman (Diffie and Hellman 1976). Αυτή ήταν η πρώτη δημοσιευμένη πρακτική μέθοδος που καθιέρωσε ένα δημόσιο κλειδί που μεταδίδεται μέσω ενός εξουσιοδοτημένου αλλά όχι ιδιωτικού καναλιού επικοινωνιών, χωρίς να χρησιμοποιείται προηγουμένως κάποιο μυστικό κλειδί.

Επίσης, η τεχνική συμφωνίας δημόσιου κλειδιού του Merkle έγινε γνωστή ως Merkle's Puzzles (Merkle and Hellman 1978), εφευρέθηκε το 1974 και δημοσιεύθηκε το 1978.

Η ιδέα των W. Diffie και M. Hellman υλοποιήθηκε ολοκληρωμένα το 1978 από τους R. Rivest, A. Shamir και L. Adleman που επινόησαν το κρυπτοσύστημα RSA (Rivest, et. al., 1978). Η ασφάλεια του RSA βασίζεται σε ένα άλλο δύσκολο μαθηματικό πρόβλημα αυτό της παραγοντοποίησης μεγάλων ακέραιων αριθμών, για το οποίο δεν υπάρχει καμία αποδοτική, δηλαδή πρακτικά γρήγορη, γενική τεχνική επίλυση. Ο RSA πραγματοποιεί κρυπτογράφηση δημόσιου κλειδιού αλλά και ψηφιακή υπογραφή δημόσιου κλειδιού.

Μετά το 1970 αναπτύχθηκε ένα μεγάλο πλήθος και ποικιλία κρυπτογραφήσεων, ψηφιακών υπογραφών, τεχνικές συμφωνίας κλειδιών και άλλες στο πεδίο της κρυπτογραφίας δημόσιου κλειδιού. Το κρυπτοσύστημα ElGamal δημιουργήθηκε το 1985 και βασίζεται στη δυσκολία επίλυσης του προβλήματος του διακριτού λογαρίθμου, όπως και το σχετικό κρυπτοσύστημα DSA που αναπτύχθηκε από την Αμερικανική Υπηρεσία Εθνικής Ασφάλειας (NSA).

Η εισαγωγή στην κρυπτογραφία ελλειπτικών καμπυλών από τον N. Koblitz και τον V. Miller ανεξάρτητα και συγχρόνως στα μέσα του 1980 οδήγησε σε νέα κρυπτοσυστήματα δημόσιου κλειδιού που βασίζονται στο πρόβλημα του διακριτικού λογαρίθμου. Παρότι είναι μαθηματικά πολυπλοκότερα, τα κρυπτοσυστήματα ελλειπτικών καμπυλών παρέχουν κλειδιά μικρότερου μεγέθους και γρηγορότερους υπολογισμούς για ισοδύναμα επίπεδα ασφαλείας.

Μία από τις σημαντικότερες συνεισφορές του δημόσιου κλειδιού κρυπτογράφησης είναι η ψηφιακή υπογραφή. Το 1991 υιοθετήθηκε το πρώτο διεθνές πρότυπο (ISO/IEC 9796). Το πρότυπο αυτό είναι βασισμένο στο σχήμα του δημόσιου κλειδιού του RSA, ενώ τα το 1994 η κυβέρνηση των Ηνωμένων Πολιτειών Αμερικής υιοθέτησε το Πρότυπο Ψηφιακής Υπογραφής (Digital Standard Signature), ένα μηχανισμό βασισμένο στο σχήμα δημόσιου κλειδιού ElGamal.
Ωστόσο σύμφωνα με τον Bellare, ο οποίος δημοσίευσε το βιβλίο του Introduction to Modern Cryptography το 2003 (Bellare and Rogaway 2003), αναφέρει ότι η κρυπτογραφία μπορεί να διαιρεθεί σε τρεις μεγάλες κατηγορίες:

- Στο πρώτο στάδιο οι διαδικασίες τις κρυπτογράφησης αφορούσαν τον τρόπο της έντυπης απεικόνισης (μελάνι και χαρτί). Έλαβαν τη μορφή αντικατάστασης και αναδιάταξης των γραμμάτων της αλφαβήτου(ενδεικτικά ο κρυπτογραφικός αλγόριθμος του Καίσαρα).
- Στο δεύτερο στάδιο αναφέρεται αυτό των κρυπτογραφικών μηχανών, ιδίως στην περίοδο του Β' παγκοσμίου πολέμου (η γερμανική μηχανή Enigma).
- Ως τρίτο στάδιο θεωρείται το σύγχρονο κρυπτογραφικό σύστημα, απόρροια της αμοιβαίας αλληλεπίδρασης των μαθηματικών και των υπολογιστών (οι υπολογιστές επέτρεψαν τη χρήση περιπλοκότερων αλγορίθμων κρυπτογράφησης και τα μαθηματικά προσέφεραν τον σχεδιασμό).

Συνοψίζοντας, η έρευνα για νέα σχήματα κρυπτογράφησης, για βελτίωση των ήδη υπαρχόντων αλγορίθμων, καθώς και για αποδείξεις της ασφάλειάς τους συνεχίζει με ραγδαίο ρυθμό. Παλαιότερες κρυπτογραφικές δομές εγκαταλείπονται και νέα προϊόντα ασφαλείας δημιουργούνται προκειμένου να καλύψουν τις ανάγκες των ευαίσθητων πληροφοριακών δεδομένων. Παρόλο την μετεξέλιξη της κρυπτογραφίας, ο στόχος παραμένει ο ίδιος: η ασφάλεια της επικοινωνίας δια μέσου ενός επισφαλούς μέσου επικοινωνίας (Pieprzyk, et. al., 2003).

2.1.3. Εισαγωγή στην κρυπτογραφία και την κρυπτανάλυση.

Η κρυπτογραφία ξεκίνησε ως η τέχνη και επιστήμη της κρυπτογράφησης, από ***«αρχαία τέχνη σε σύγχρονη επιστήμη»*** *όπως αναφέρει και ο U. Maurer (Maurer 2001)* και αφορούσε στη δημιουργία τεχνικών μεταβολής ενός μηνύματος με τέτοιο τρόπο ώστε να διατηρείται μυστικό το περιεχόμενό του.

Σήμερα η κρυπτογραφία είναι η επιστήμη που μελετά μαθηματικές μεθόδους σχετιζόμενες με διάφορους τομείς της ασφάλειας πληροφορίας, όπως η εμπιστευτικότητα, η ακεραιότητα των δεδομένων, η πιστοποίηση, τεχνικές οι οποίες θα αναλυθούν στην συνέχεια του κεφαλαίου.

2.1.4. Ορολογία

Αρχικό κείμενο (plaintext) είναι το μήνυμα το οποίο αποτελεί την είσοδο σε μία διεργασία κρυπτογράφησης. Είναι σε μορφή που μπορεί να γίνει κατανοητή από ένα άτομο (π.χ ένα έγγραφο) είτε από έναν υπολογιστή (π.χ εκτελέσιμος κώδικας) (van Oorschot and Wiener 2006).

Κλειδί (key) είναι ένας αριθμός αρκετών bit που χρησιμοποιείται ως είσοδος στην συνάρτηση κρυπτογράφησης (Rivest et. al., 1978).

Η **Κρυπτογραφία (Cryptography)** μελετά τρόπους με τους οποίους μπορούμε να μετασχηματίσουμε ένα μήνυμα σε ακατάληπτη μορφή (Κάτος και Στεφανίδης 2003). Η κρυπτογραφία ασχολείται με την επικοινωνία δεδομένης της ύπαρξης του αντιπάλου. Αυτή η ύπαρξη αντιπάλου σε κάποια επικοινωνία αποτελεί και την βασική αιτία ύπαρξης και εφαρμογής της κρυπτογραφίας (C. Kaufman, Perlman, και Speciner 2002; Delfs and Knebl 2007; Bishop 2003)

Αντίπαλος (adversary) θεωρείται οποιαδήποτε οντότητα, εκτός του αποστολέα και του παραλήπτη, η οποία προσπαθεί να ανατρέψει την ασφάλεια του συστήματος επικοινωνίας μεταξύ τους. Ένας αντίπαλος σε πολλές περιπτώσεις μπορεί να προσπαθήσει να υποδυθεί τον έγκυρο αποστολέα ή παραλήπτη (Λάσκαρη 2010).

Η **Κρυπτογράφηση (Encryption)** είναι η μέθοδος που μετασχηματίζει τα αρχικά δεδομένα, το απλό κείμενο όπως έχει αναφερθεί παραπάνω, σε μία μορφή που εμφανίζεται να είναι τυχαία και δυσανάγνωστη.

Η **Κρυπτανάλυση (cryptanalysis)** είναι η επιστήμη που ασχολείται με την διαδικασία ανεύρεσης του απλού κειμένου από ένα κρυπτογραφημένο μήνυμα, χωρίς να έχουν πρόσβαση στο κλειδί της κρυπτογράφησης (Γιαννακόπουλος 2007). Επίσης ορίζεται ως η μελέτη των μαθηματικών τεχνικών, οι οποίες επιχειρούν την αναίρεση των τεχνικών της Κρυπτογραφίας και γενικότερα την αναίρεση της αφαλούς μεταδόσεως πληροφοριών (Λάσκαρη 2010). Η κρυπτανάλυση βασίζεται πέραν των μαθηματικών και στο εμπειρικό γεγονός ότι,

στην πράξη, ο κρυπταναλυτής έχει στη διάθεσή του πάρα πολύ μεγάλο αριθμό κρυπτογραφημένων κειμένων, τα οποία κρυπτογραφήθηκαν με τον ίδιο τρόπο. Ενώ επίσης θεωρείται δεδομένο το γεγονός ότι ο κρυπταναλυτής γνωρίζει πλήρως τη διαδικασία κρυπτογράφησης-αποκρυπτογράφησης που χρησιμοποιήθηκε. Οι διαδικασίες κρυπτανάλυσης ονομάζονται επιθέσεις (attacks).

Σκοπός του κρυπταναλυτή είναι να αποκτήσει:

1. Ένα μέρος από κάποιο αρχικό καθαρό κείμενο ή ολόκληρο.
2. Ένα μέρος από κάποιο κρυπτογραφημένο κείμενο ή ολόκληρο.
3. Συνδυασμό των προηγούμενων.
4. Συνδυασμό των προηγούμενων από διαφορετικά μηνύματα.
5. Γνώση παραγωγής ή απόκτησης των κλειδιών κρυπτογράφησης.

Κρυπτογραφημένο κείμενο ή κρυπτοκείμενο (ciphertext) είναι το αποτέλεσμα της εφαρμογής ενός κρυπτογραφικού αλγορίθμου πάνω στο αρχικό κείμενο.

Η μετατροπή του κρυπτοκειμένου σε απλό κείμενο ονομάζεται **αποκρυπτογράφηση (decryption).**

Το σύστημα το οποίο παρέχει την κρυπτογράφηση και την αποκρυπτογράφηση αναφέρεται ως **κρυπτογραφικό σύστημα ή κρυπτοσύστημα (cryptosystem)**.

Κρυπτογραφικός αλγόριθμος είναι η συνάρτηση ή το σύνολο των κανόνων, στοιχείων και βημάτων που καθορίζουν την κρυπτογράφηση και αποκρυπτογράφηση. Ορισμένες φορές καλείται και **κωδικοποιητής (cipher)** (Ζορκάδης 2002). Είναι η μέθοδος μετασχηματισμού δεδομένων σε μία μορφή που να μην επιτρέπει την αποκάλυψη των περιεχομένων τους από μη εξουσιοδοτημένα μέλη. Κατά κανόνα ο κρυπτογραφικός αλγόριθμος είναι μία πολύπλοκη μαθηματική συνάρτηση. Δέχεται ως είσοδο ένα αρχικό μήνυμα (plaintext) και δίνει στην έξοδο ένα τροποποιημένο μήνυμα (ciphertext).

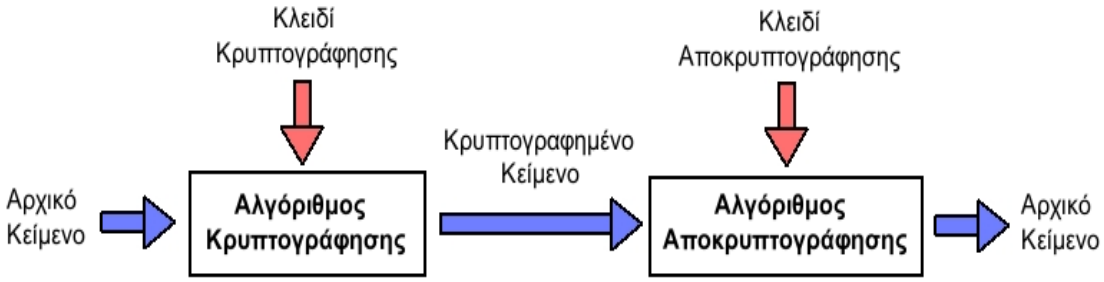
Οι περισσότερες μέθοδοι κρυπτογράφησης χρησιμοποιούν μία *μυστική αξία (secret value)* την αποκαλούμενη **κλειδί**, που λειτουργεί με τον αλγόριθμο για να κρυπτογραφήσει και να αποκρυπτογραφήσει το κείμενο. Η ύπαρξη του κλειδιού είναι αυτή η οποία προσδιορίζει την διαφορά της κρυπτογράφησης από την κωδικοποίηση (encoding).

Κρυπτογραφικά πρωτόκολλα καλούνται τα πρωτόκολλα τα οποία χρησιμοποιούν κρυπτογραφικούς αλγορίθμους (Delfs and Knebl 2007; Washington and Trappe 2002; Stallings 2003).

Παράδειγμα:

Προκειμένου να γίνουν κατανοητές όλες αυτές οι έννοιες της κρυπτογραφίας, αναλύεται ένα απλό πρόβλημα μέσω του οποίου μπορούν να συνδυαστούν αυτές οι βασικές έννοιες αλλά και να περιγραφεί η βασική λειτουργία της διαδικασίας της κρυπτογράφησης. Στο πρόβλημα λοιπόν, αυτό θεωρείται ότι υπάρχει ο αποστολέας ο οποίος θέλει να στείλει ένα μήνυμα, το αρχικό κείμενο, σε έναν παραλήπτη. Υπάρχει ένας αντίπαλος ο οποίος θα προσπαθήσει να ανατρέψει την ασφάλεια του συστήματος επικοινωνίας. Γι' αυτό λοιπόν το αρχικό, καθαρό μήνυμα που θέλει να μεταβιβάσει ο παραλήπτης, υπόκειται την διαδικασία της κρυπτογράφησης, μέσω της οποίας θα διατηρηθεί κρυφό το περιεχόμενό του. Το νέο αυτό κείμενο είναι το κρυπτοκείμενο.

Η διαδικασία της κρυπτογράφησης και της αποκρυπτογράφησης φαίνεται στο παρακάτω σχήμα.



Σχήμα 2.1: Η διαδικασία της κρυπτογράφησης και της αποκρυπτογράφησης.

Η κρυπτογράφηση και αποκρυπτογράφηση ενός μηνύματος γίνεται με τη βοήθεια ενός αλγορίθμου κρυπτογράφησης (cipher) και ενός κλειδιού κρυπτογράφησης (key). Συνήθως ο αλγόριθμος κρυπτογράφησης είναι γνωστός, οπότε η εμπιστευτικότητα του κρυπτογραφημένου μηνύματος που μεταδίδεται βασίζεται ως επί το πλείστον στην μυστικότητα του κλειδιού κρυπτογράφησης. Το μέγεθος του κλειδιού κρυπτογράφησης μετριέται σε bit. Ένας αλγόριθμος περιέχει ένα εύρος κλειδιών (keyspace), το οποίο είναι μια σειρά τιμών που μπορεί να χρησιμοποιηθούν για να κατασκευάσει το κλειδί. Το κλειδί αποτελείται από τυχαίες τιμές μέσα στα όρια διακύμανσης του εύρους. Γενικά, ισχύει ο εξής κανόνας: **όσο μεγαλύτερο είναι το κλειδί κρυπτογράφησης, τόσο πιο πολλές διαθέσιμες τιμές μπορούν να χρησιμοποιηθούν για να αντιπροσωπεύσουν τα διαφορετικά κλειδιά και τόσο περισσότερο τυχαία είναι τα κλειδιά και άρα**

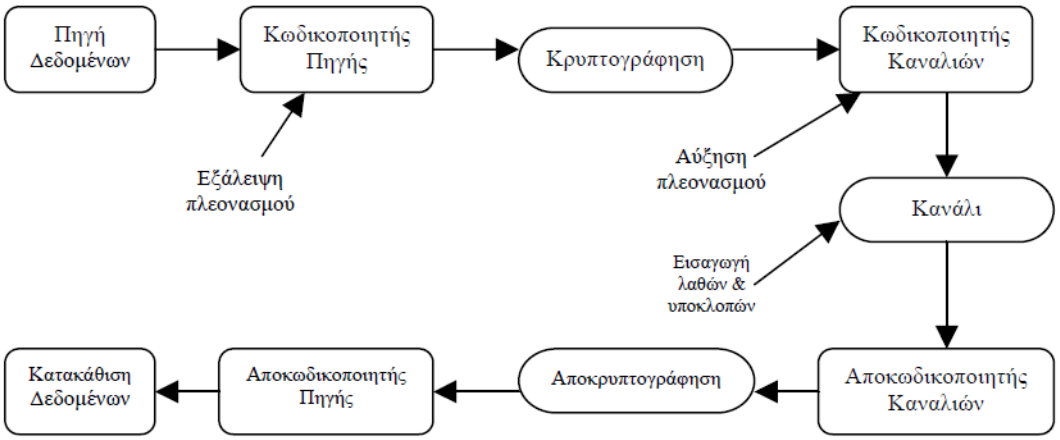
δυσκολότερα πρόκειται να αποκρυπτογραφηθεί το κρυπτογραφημένο μήνυμα από επίδοξους εισβολείς. Διαφορετικοί αλγόριθμοι κρυπτογράφησης απαιτούν διαφορετικά μήκη κλειδιών για να πετύχουν το ίδιο επίπεδο ανθεκτικότητας κρυπτογράφησης. Πολλοί αλγόριθμοι είναι δημόσια γνωστοί και δεν αποτελούν το μυστικό μέρος της διαδικασίας κρυπτογράφησης.

2.2. ΚΡΥΠΤΟΓΡΑΦΙΑ ΚΑΙ ΚΩΔΙΚΟΠΟΙΗΣΗ

Υπάρχουν τρεις βασικές μορφές κωδικοποίησης στα σύγχρονα συστήματα επικοινωνιών. Αυτά είναι: (Ζορκάδης 2002; Stallings 2003; Stinson 2006)

1. **κωδικοποιητής της πηγής** (source coding),
2. **κωδικοποιητής λάθους** (error coding -επίσης αποκαλείται και κωδικοποιητής καναλιών (channel coding) και
3. **κρυπτογράφηση** (encryption).

Από θεωρητική και πρακτική πληροφοριακή πτυχή, οι τρεις μορφές κωδικοποίησης πρέπει να εφαρμοστούν ως εξής (σχήμα2.2),(Stallings 2003):



Σχήμα 2.2: Η κωδικοποίηση στα ψηφιακά συστήματα επικοινωνίας.

Κωδικοποιητής πηγής (συμπίεση στοιχείων): Τα περισσότερα δεδομένα, όπως το κείμενο, εμπεριέχουν την έννοια του πλεονασμού. Αυτό σημαίνει τυποποιημένη απεικόνιση του μηνύματος, π.χ το αγγλικό κείμενο χρησιμοποιεί περισσότερα bitsαπό τα απαραίτητα για να αντιπροσωπεύσει τον πλεονασμό και επομένως μειώνουν την έκταση μηνυμάτων.

Κρυπτογράφηση: Ο στόχος της κρυπτογράφηση είναι να μετασχηματίζει το περιεχόμενο ενός μηνύματος σε ακατάληπτη μορφή. Μόνο ο κάτοχος των κρυπτογραφικών κλειδιών πρέπει να είναι σε θέση να ανακτήσει το αρχικό περιεχόμενο. Η κρυπτογράφηση μπορεί να αντιμετωπισθεί ως μορφή κωδικοποίησης.

Κωδικοποίηση καναλιών (κωδικοποίηση λάθους): Ο σκοπός των κωδίκων καναλιών είναι να καταστούν τα δεδομένα εύρωστα ενάντια στα λάθη που προκύπτουν κατά τη διάρκεια της μετάδοσης μέσα στο κανάλι.

2.2.1. Κρυπτογραφικοί αλγόριθμοι.

Οι κρυπτογραφικοί αλγόριθμοι χρησιμοποιούν, κατά κανόνα (κρυπτογραφικά) **κλειδιά** (keys), η τιμή των οποίων επηρεάζει την κρυπτογράφηση και την αποκρυπτογράφηση. Το σύνολο των δυνατών τιμών των κλειδιών λέγεται **πεδίο τιμών** (keyspace).

Υπάρχουν δύο κατηγορίες κρυπτογραφικών αλγορίθμων, και κατά συνέπεια συστημάτων: **οι συμμετρικοί αλγόριθμοι και οι ασύμμετροι** αλγόριθμοι (Ζορκάδης 2002; Λάσκαρη 2010; Stinson 2006; Stallings 2003).

Οι συμμετρικοί αλγόριθμοι χρησιμοποιούν το ίδιο κλειδί για την κρυπτογράφηση και την αποκρυπτογράφηση, και για τον λόγο αυτό καλούνται επίσης **αλγόριθμοι μυστικού κλειδιού** ή **αλγόριθμοι μονού κλειδιού**.

Οι ασύμμετροι αλγόριθμοι χρησιμοποιούν ένα ζεύγος κρυπτογραφικών κλειδιών: το δημόσιο κλειδί και το ιδιωτικό κλειδί. Είναι δύο ξεχωριστά κλειδιά τα οποία όμως σχετίζονται μεταξύ τους. Τα κλειδιά που ανήκουν στο ζεύγος αυτό έχουν τη σημαντική ιδιότητα ότι είναι πρακτικά αδύνατος ο υπολογισμός του ενός κλειδιού γνωρίζοντας το άλλο.

Η ανάλυση και των δύο κατηγοριών θα γίνει στην συνέχεια του κεφαλαίο, αφού γίνει μια αναφορά στην αναγκαιότητα της κρυπτογραφίας και τους λόγους ύπαρξης αυτών.

2.2.2. Αναγκαιότητα της Κρυπτογραφίας.

Προκειμένου να επιτευχθούν οι στόχοι ασφαλείας, όπως η μυστικότητα και η αυθεντικότητα, χαρακτηριστικά τα οποία θα αναλυθούν στην συνέχεια του

κεφαλαίου, το κρυπτοσύστημα που χρησιμοποιούμε παρέχει στον αποστολέα και στον αποδέκτη ένα πρωτόκολλο. Το πρωτόκολλο είναι μια συλλογή προγραμμάτων (αλγόριθμοι, λογισμικό), ένα για κάθε ενδιαφερόμενο μέρος.
Σύμφωνα με τον Bruce Schneiner ο οποίος το 1996 δημοσίευσε το βιβλίο του "Applied Cryptography" (Schneier 2007) **αναφέρεται ότι η κρυπτογραφία αφορά την κατασκευή και την ανάλυση των πρωτοκόλλων ασφαλείας.** Το πρωτόκολλο είναι αυτό το οποίο αποδίδει τους κανόνες λειτουργίας και συμπεριφοράς σε κάθε συμβαλλόμενο μέρος. Το πρωτόκολλο είναι ουσιαστικά ένα κατανεμημένο πρόγραμμα. Συνεπώς τίθεται το ζήτημα της ανάλυσης αυτών των πρωτοκόλλων. *Το πρώτο βήμα* είναι να γίνουν αντιληπτές οι απειλές και οι στόχοι για το πρόβλημα της ασφαλείας, ώστε να βρεθεί μία λύση του πρωτοκόλλου αυτού.
Στην κρυπτογραφία υπάρχουν κανόνες, ο σημαντικότερος εκ των οποίων είναι η επίτευξη να νικηθεί ο αντίπαλος με τη βοήθεια των πρωτοκόλλων. *Ο δεύτερος κανόνας* είναι ότι τα πρωτόκολλα πρέπει να είναι δημοσίως γνωστά. Ο κανόνας ο οποίος παραμένει μυστικός ενσωματώνεται στα κλειδιά.

2.2.3. Απειλές και επιπτώσεις παραβίασης της ασφάλειας.

Οι διάφορες απόπειρες παραβίασης της ασφάλειας ακούσιες ή εκ προθέσεως, και οι αντίστοιχες επιπτώσεις διακρίνονται στις εξής κατηγορίες: διακοπής ή άρνησης υπηρεσίας, υποκλοπής, παραποίησης, πειρατείας και αμφισβήτησης.
Όπως έχει αναφερθεί στην παράγραφο 2.1.3 η κρυπτανάλυση είναι η μελέτη των μαθηματικών μεθόδων προκειμένου να ανατρέψει την ασφάλεια των κρυπτοσυστημάτων και των υπηρεσιών της ασφαλούς επικοινωνίας γενικότερα. **Θεμελιώδης υπόθεση της κρυπτανάλυσης είναι η υπόθεση *Kercκhoff* (Delfs and Knebl 2007), σύμφωνα με την οποία η μυστικότητα πρέπει ολοκληρωτικά να βασίζεται στο κλειδί κρυπτογράφησης.** Ουσιαστικά η κρυπτανάλυση στοχεύει στη αποκάλυψη του πρωτότυπου κειμένου χωρίς να υπάρχει πρόσβαση στο κλειδί και η επιτυχής κρυπτανάλυση μπορεί να αποκαλύψει είτε το αρχικό μήνυμα είτε το κλειδί που χρησιμοποιείται για την κρυπτογράφησή του, μέσω των αδύναμων σημείων του κρυπτοσυστήματος.
Μία προσπάθεια κρυπτανάλυσης καλείται επίθεση (attack). Υπάρχουν διάφοροι γενικοί τύποι κρυπταναλυτικών επιθέσεων, οι πιο σημαντικοί τύποι επιθέσεων που

εφαρμόζονται σε αλγορίθμους είναι οι παρακάτω (Ζορκάδης 2002; Cramer and Shoup 1998; Rackoff and Simon 1992; Jakobsen and Knudsen 1997):

1. Επίθεση μόνο κρυπτογραφήματος (Ciphertext-only attack).
2. Επίθεση γνωστού πρωτοτύπου κειμένου (Known-plaintext attack).
3. Επίθεση επιλεγμένου πρωτοτύπου κειμένου (Chosen-plaintext attack).
4. Επίθεση μεταβαλλόμενου πρωτοτύπου κειμένου (Adaptive-chosen-plaintext attack).
5. Επίθεση επιλεγμένου κρυπτογραφήματος (Chosen ciphertext attack).

2.2.4. Κρυπτογραφικές αρχές.

Παρά την ύπαρξη διαφόρων ειδών κρυπτογραφικών συστημάτων έχει μελετηθεί ότι όλα, ανεξάρτητα από τον τρόπο λειτουργίας τους, ακολουθούνε δύο θεμελιώδεις αρχές οι οποίες και θα αναλυθούν.

Η πρώτη αρχή που διέπει τα κρυπτογραφικά συστήματα είναι ο **"πλεονασμός" (redundancy)** (Rivest et. al., 1978; Stallings 2003)**.** Τα κρυπτογραφημένα μηνύματα πρέπει να παρέχουν πληροφορίες οι οποίες δεν απαιτούνται για την κατανόηση του μηνύματος. Με αυτόν τον τρόπο, γίνεται δύσκολο στους εισβολείς να στείλουν τυχαίες άχρηστες πληροφορίες, με σκοπό να λαμβάνονται ως έγκυρα μηνύματα τα οποία έχουν αλλοιωθεί. Ωστόσο, η προσθήκη αυτή στα μηνύματα τα καθιστά πιο ευάλωτα σε τυχόν επιθέσεις. Ο πλεονασμός δεν πρέπει ποτέ να έχει την μορφή *n* μηδενικών στη αρχή ή στο τέλος ενός μηνύματος, επειδή η χρήση τέτοιων μηνυμάτων σε κάποιους κρυπτογραφικούς αλγορίθμους δίνει πιο προβλέψιμα αποτελέσματα, γεγονός που κάνει ευκολότερη την παραβίασή τους.

Μία δεύτερη κρυπτογραφική αρχή είναι η λεγόμενη **"φρεσκάδα" (freshness)** (Rivest et. al., 1978; Stallings 2003) την οποία θα πρέπει να έχουν τα λαμβανόμενα μηνύματα, δηλαδή να έχουν αποσταλεί πάρα πολύ πρόσφατα. Με αυτόν τον τρόπο αποτρέπεται η αναπαραγωγή παλιών μηνυμάτων από τους εισβολείς. Αυτή η αρχή γενικότερα εκφράζει την ανάγκη ύπαρξης κάποιας μεθόδου ματαίωσης των επιθέσεων αναπαραγωγής. Ένα τέτοιο μέτρο είναι να περιλαμβάνεται σε κάθε μήνυμα μια χρονοσφραγίδα που θα είναι έγκυρη για πάρα πολύ μικρό χρονικό διάστημα, για παράδειγμα για δέκα δευτερόλεπτα. Έτσι, ο παραλήπτης μπορεί να διατηρεί τα μηνύματα για το συγκεκριμένο χρονικό όριο ώστε να μπορεί να τα συγκρίνει με νεοεισερχόμενα και να φιλτράρει τα αντίγραφα.

Τα μηνύματα εκείνα τα οποία θα είναι παλιότερα από τα δέκα δευτερόλεπτα θα απορρίπτονται.

2.2.5. Βασικοί στόχοι της ασφάλειας.

Σύμφωνα με το «Εγχειρίδιο Εφαρμοσμένης Κρυπτογραφίας» (Menezes, et.al., 1997), υπάρχουν αρκετοί αντικειμενικοί σκοποί για την ασφάλεια των πληροφοριών εκ των οποίων δίνεται βαρύτητα στους ακόλουθους τέσσερις, οι οποίοι δημιουργούν ένα πλαίσιο βάση των οποίων θα εξαχθούν οι υπόλοιποι: 1) Εμπιστευτικότητα ή μυστικότητα, 2) ακεραιότητα δεδομένων, 3) πιστοποίηση αυθεντικότητας, 4) μη-απάρνηση.

1. *Εμπιστευτικότητα* είναι μια υπηρεσία που χρησιμοποιείται για να κρατά το περιεχόμενο της πληροφορίας μακριά από όλους εκτός από εκείνους που είναι εξουσιοδοτημένοι να το έχουν. Η *εχεμύθεια* είναι ένας όρος συνώνυμος με την εμπιστευτικότητα και τη μυστικότητα. Υπάρχουν διάφορες προσεγγίσεις για την εξασφάλιση της εμπιστευτικότητας, που κυμαίνονται από τη φυσική προστασία μέχρι μαθηματικούς αλγορίθμους οι οποίοι αποδίδουν τα δεδομένα σε κατάλληλη μορφή.
2. *Ακεραιότητα των δεδομένων* είναι μία υπηρεσία που απευθύνεται στη μη εξουσιοδοτημένη μεταβολή των δεδομένων. Για τη διασφάλιση της ακεραιότητας των δεδομένων κάποιος πρέπει να έχει τη δυνατότητα να ανιχνεύει το χειρισμό των δεδομένων από μη εξουσιοδοτημένα μέλη. Ο χειρισμός των δεδομένων περιλαμβάνει τέτοιες ενέργειες όπως η εισαγωγή, η διαγραφή και η αντικατάσταση.
3. *Πιστοποίηση αυθεντικότητας* είναι μια υπηρεσία που σχετίζεται με την ταυτοποίηση. Αυτή η λειτουργία εφαρμόζεται στις δύο οντότητες και την ίδια πληροφορία. Δύο μέλη που εισέρχονται σε μια επικοινωνία θα πρέπει να ταυτοποιήσουν το ένα το άλλο. Στην πληροφορία που μεταβιβάζεται μέσω ενός καναλιού, θα πρέπει να πιστοποιείται η αυθεντικότητά της ως προς την πηγή προέλευσης, η ημερομηνία προέλευσης, το περιεχόμενο των δεδομένων, η ώρα αποστολής, κτλ. Γι' αυτούς τους λόγους αυτή η πλευρά της κρυπτογραφίας συνήθως υποδιαιρείται σε δύο κύριες κλάσεις: την *πιστοποίηση αυθεντικότητας της οντότητας* και *πιστοποίηση της πηγής δεδομένων*. Η πιστοποίηση αυθεντικότητας της πηγής των δεδομένων

έμμεσα παρέχει την ακεραιότητα των δεδομένων (γιατί αν ένα μήνυμα είναι τροποποιημένο, τότε έχει αλλάξει η πηγή).

4. *Μη-απάρνηση* είναι μια υπηρεσία η οποία αποτρέπει μια οντότητα από το να αρνηθεί προηγούμενες δεσμεύσεις ή ενέργειες. Όταν προκύψουν αμφισβητήσεις που οφείλονται στο γεγονός ότι μια οντότητα αρνείται ότι είχαν γίνει ορισμένες ενέργειες, είναι απαραίτητο ένα μέσον προκειμένου να αποσαφηνίσει την κατάσταση. Παραδείγματος χάρη, μια οντότητα μπορεί να δώσει εξουσιοδότηση για την αγορά ενός αγαθού από μια άλλη οντότητα και αργότερα να αρνηθεί ότι είχε παραχωρηθεί τέτοια εξουσιοδότηση. Είναι απαραίτητη μια διαδικασία που εμπλέκει ένα έμπιστο τρίτο μέλος για να άρει την αμφισβήτηση.

Γενικά, ένας από τους βασικούς στόχους της κρυπτογραφίας είναι να εξασφαλιστεί η καλύτερη δυνατή ικανοποίηση των τεσσάρων προηγούμενων υπηρεσιών ασφαλείας τόσο στη θεωρία όσο και στην πράξη. Αντικειμενικός της σκοπός είναι να ανακαλύψει και να αποτρέψει οποιαδήποτε προσπάθεια εξαπάτησης ή κακόβουλη ενέργεια (Goldreich 2001; Zlotkin et. al., 1994; Hankerson, et. al., 2004).

2.3. ΔΥΝΑΜΗ ΚΡΥΠΤΟΓΡΑΦΙΚΟΥ ΣΥΣΤΗΜΑΤΟΣ

2.3.1. Αρχή του Kerckhoff.

Ένα θεμελιώδες κριτήριο στην αντικειμενική μέτρηση της δύναμης ενός κρυπτοσυστήματος είναι η αρχή Kerckhoff:
«*Η ασφάλεια ενός κρυπτοσυστήματος δεν εξαρτάται από την μυστικότητα του αλγορίθμου κρυπτογράφησης. Η ασφάλεια του κρυπτοσυστήματος εξαρτάται μόνον από το να διατηρείται μυστικό το κλειδί*» (Delfs and Knebl 2007).

2.3.2. Βασικές αρχές σχεδιασμού κρυπτογραφημάτων ομάδας (block ciphers).

2.3.2.1. Τα μέτρα του Shannon.

Ο Shannon, ο θεμελιωτής της θεωρίας της πληροφορίας διατύπωσε το 1949 ένα σύνολο από μέτρα τα οποία χαρακτηρίζουν ένα ορθά σχεδιασμένο αλγόριθμο κρυπτογράφησης (Shannon 1949):

1. Βαθμός απαιτούμενης κρυπτογραφικής ασφάλειας. Το μέτρο αυτό αφορά το κέρδος του αντιπάλου σε πληροφορία, όταν παρατηρεί το κρυπτοκείμενο.
2. Μήκος του κλειδιού. Η ευκολία χειρισμού του κλειδιού εξαρτάται από το μήκος του.
3. Πρακτική εκτέλεση της κρυπτογράφησης και της αποκρυπτογράφησης. Η προσπάθεια που απαιτείται για την κρυπτογράφηση και την αποκρυπτογράφηση, σε χρόνο ή λειτουργίες.
4. Διόγκωση του κρυπτοκειμένου. Είναι επιθυμητό το κρυπτοκείμενο να έχει το ίδιο μήκος (ή συγκρίσιμου μεγέθους) με το απλό κείμενο.
5. Διάδοση των σφαλμάτων κρυπτογράφησης. Είναι επιθυμητό ένα σφάλμα κατά την κρυπτογράφηση να επηρεάζει σε όσο το δυνατό λιγότερο βαθμό την αποκρυπτογράφηση.

Η ύπαρξη των μέτρων σε ένα κρυπτοκείμενο είναι υποχρεωτική, αλλά συγχρόνως και αντιφατική, με αποτέλεσμα να μην υπάρχει στην πραγματικότητα κρυπτοσύστημα το οποίο ικανοποιεί όλα τα μέτρα στο μέγιστό τους. Για παράδειγμα, πλήρης έλλειψη του πρώτου μέτρου σημαίνει ότι ο αντίπαλος μπορεί να ανακτήσει πλήρως το απλό κείμενο. Η πλήρης έλλειψη του τρίτου και τέταρτου μέτρου επιτρέπει κρυπτό-συστήματα που μπορούν να μεγιστοποιούν όλα τα άλλα μέτρα, αλλά σε περίπτωση σφάλματος κατά την κρυπτογράφηση, η ανάκτηση του απλού κειμένου θα ήταν αδύνατη, ακόμη και κάποιο τμήμα αυτού.

2.3.2.2. Σύγχυση (confusion) και Διάχυση (diffusion).

Δύο ιδιότητες που χρησιμοποιούνται στην σωστή σχεδίαση ενός κρυπταλγορίθμου είναι η σύγχυση (confusion) και η διάχυση (diffusion), (Shannon 1949).

Σύγχυση είναι η ικανότητα του αλγορίθμου κρυπτογράφησης όπου ο αντίπαλος δεν είναι σε θέση να προβλέψει ποιες μεταβολές θα συμβούν στο κρυπτοκείμενο, δεδομένης μιας μεταβολής στο απλό κείμενο. Δηλαδή, ένας αλγόριθμος έχει υψηλή σύγχυση όταν η σχέση μεταξύ του απλού κειμένου και του κρυπτοκειμένου

είναι αρκετά πολύπλοκη, ώστε να χρειάζεται ο αντίπαλος να ξοδέψει σημαντικό χρόνο προκειμένου να τις προσδιορίσει.

Διάχυση είναι η ικανότητα του αλγορίθμου κρυπτογράφησης όπου ένα τμήμα του απλού κειμένου να έχει την ευκαιρία να επηρεάζει όσο το δυνατόν περισσότερα τμήματα του κρυπτοκειμένου. Ένας αλγόριθμος έχει υψηλή διάχυση όταν ένα στοιχειώδες τμήμα του απλού κειμένου έχει την δυνατότητα να επηρεάσει όλα τα τμήματα του κρυπτοκειμένου ανεξαρτήτως της τοποθεσίας του τμήματος αυτού στο απλό κείμενο.

Έστω ένα απλό κείμενο το οποίο αντιστοιχεί σε ένα κρυπτοκείμενο μέσω ενός κρυπταλγορίθμου. Εάν αντικαταστήσουμε ένα σύμβολο του απλού κειμένου και κρυπτογραφήσουμε το νέο απλό κείμενο, τότε για ένα κρυπταλγόριθμο με υψηλή διάχυση, ο αντίπαλος δεν θα μπορεί να προβλέψει ποια σύμβολα του κρυπτοκειμένου θα μεταβληθούν ή γενικότερα θα επηρεαστούν.

Συμπερασματικά για την δύναμη της μεθόδου κρυπτογράφησης μπορεί να ειπωθεί ότι αποτελεί ένα συνδυασμό

- του αλγορίθμου,
- της μυστικότητα του κλειδιού,
- του μήκος του κλειδιού, και
- του τρόπο που λειτουργούν από κοινού μαζί.

Γενικότερα όμως, όταν γίνεται αναφορά στον βαθμό της ισχύος της κρυπτογράφησης, συνήθως εννοείται ο βαθμό δυσκολίας για τον υπολογισμό του αλγορίθμου ή του κλειδιού, όταν ένα από τα δύο δεν δημοσιοποιείται.

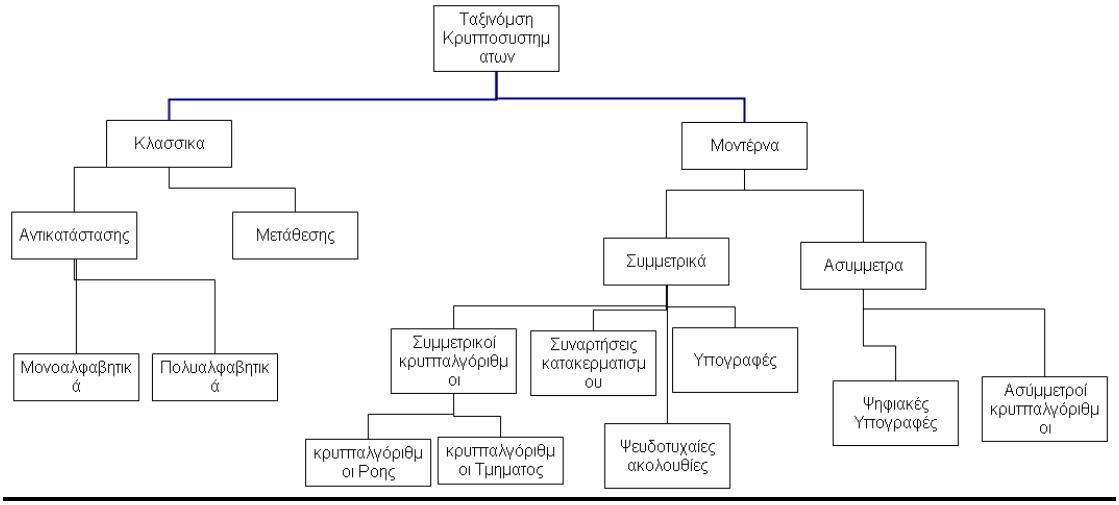
Η δύναμη ενός κρυπτοσυστήματος να αντιστέκεται στις επιθέσεις του αντίπαλου είναι ένα αντικείμενο το οποίο μπορεί να εξεταστεί από πολλές πλευρές. Η ανάγκη καθορισμού αντικειμενικών μέτρων για τη μέτρηση της κρυπτογραφικής δύναμης είχε ως αποτέλεσμα τη δημιουργία διάφορων μαθηματικών μοντέλων. Τα μοντέλα αυτά είναι (Beaver et.al., 1991; Cachin 2005):

- **Ασφάλεια άνευ όρων** (unconditionally secure). Ένα κρυπτοσύστημα είναι άνευ όρων ασφαλές όταν το κρυπτοκείμενο δεν δίνει καμία πληροφορία στον αντίπαλο σχετικά με το απλό κείμενο. Η υπόθεση απαιτεί ότι ο αντίπαλος έχει άπειρη υπολογιστική ισχύ στη διάθεσή του.

- **Υπολογιστική ασφάλεια** (computationally secure). Ένα κρυπτοσύστημα είναι υπολογιστικά ασφαλές, όταν προκειμένου να το σπάσει ο αντίπαλος απαιτείται υπολογιστική ισχύς πέραν των δυνατοτήτων του.
- **Ασφάλεια θεωρητικής πολυπλοκότητας** (complexity theoretic). Θεωρείται ότι ο αντίπαλος μπορεί να πραγματοποιήσει επίθεση στο κρυπτοσύστημα η οποία απαιτεί πολυωνυμική υπολογιστική ισχύ. Δηλαδή, οι παράμετροι ασφαλείας του κρυπτοσυστήματος μπορούν να εκφραστούν πολυωνυμικά ως προς το χώρο και το χρόνο.
- **Αποδείξιμη ασφάλεια** (provable security). Ένα κρυπτοσύστημα είναι αποδείξιμα ασφαλές όταν μπορούμε να αποδείξουμε ότι η ασφάλειά του είναι ισοδύναμη κάποιου γνωστού και καλά μελετημένου προβλήματος που θεωρείται "δύσκολο".

2.4. Είδη κρυπτογράφησης.

Τα κρυπτοσυστήματα χωρίζονται σε δύο μεγάλες κατηγορίες: **τα Κλασικά Κρυπτοσυστήματα και τα Μοντέρνα Κρυπτοσυστήματα.** Στην παρακάτω εικόνα παρουσιάζεται αναλυτικά ο διαχωρισμός των δύο αυτών μεγάλων κατηγοριών των κρυπτοσυστημάτων (Gordon and Jeffrey 2005; Kaufman and Matyas 1998; Burnett and Paine 2001).



Σχήμα 2.3:Είδη κρυπτοσυστημάτων.

2.4.1. Κλασικά Κρυπτοσυστήματα.

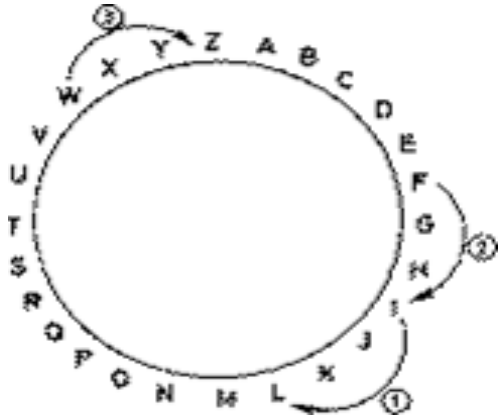
Κλασικά κρυπτοσυστήματα αποκαλούνται συνήθως τα κρυπτοσυστήματα αντικατάστασης και τα κρυπτοσυστήματα αναδιάταξης ή μετάθεσης.

2.4.1.1. Κρυπτογράφημα Αντικατάστασης.

Ένα κρυπτογράφημα αντικατάστασης είναι αυτό όπου τα γράμματα του σχεδίου του κειμένου αντικαθίσταται από άλλα γράμματα ή από νούμερα ή από σύμβολα. Αν το σχέδιο κειμένου αποτελείται από ακολουθία bits, τότε η αντικατάσταση περιλαμβάνει αντικατάσταση προτύπων bit του κειμένου από πρότυπα bit κρυπτοκειμένου.

2.4.1.1.1. Μονοαλφαβητικό κρυπτογράφημα.

Η πιο γνωστή χρήση του κρυπτογραφήματος ήταν από τον Ιούλιο Καίσαρα. Αυτό είναι γνωστό σαν μονοαλφαβητικό κρυπτογράφημα. Στην εικόνα 2.5 παρουσιάζεται ο τρόπος ολίσθησης των γραμμάτων.



Εικόνα 2.5: Μονοαλφαβητικό κρυπτογράφημα του Καίσαρα.

Στον παρακάτω σχήμα δίνεται ένα κανονικό σχέδιο κειμένου και ακριβώς από κάτω το κρυπτοκείμενο:

Κείμενο	ABCDEFGHIJKLMNOPQRSTUVWXYZ
Κρυπτοκείμενο	DEFGHIJKLMNOPQRSTUVWXYZABC

Σχήμα 2.4 : Αντιστοίχιση κρυπτοκειμένου σε κείμενο.

Τα μεταβαλλόμενα νούμερα που χρησιμοποιούνται για να μετατρέψουν ένα σχέδιο κειμένου σε κρυπτοκείμενο αναφέρονται συνήθως ως κλειδί. Το μειονέκτημα ενός τέτοιου συστήματος είναι ότι μπορεί να δεχθεί μεγάλη επίθεση απλώς δοκιμάζοντας την λειτουργικότητα της χρησιμοποιούμενης γλώσσας όταν η φύση του σχεδίου του κειμένου είναι γνωστή. **Η επίθεση αυτή είναι γνωστή και σαν συχνότητα ανάλυσης.** Μία τέτοια επίθεση συχνότητας ανάλυσης μπορεί να γίνει μόνο με απλή εκτέλεση ενός προγράμματος το οποίο μας δείχνει την συγγενική συχνότητα των γραμμάτων του μηνύματος.

2.4.1.1.2. Πολυαλφαβητικό Κρυπτογράφημα.

Πολυαλφαβητικό σύστημα αντικατάστασης ονομάζεται κάθε τεχνική που επιτρέπει διαφορετικά κρυπτογραφήματα να αναπαριστούν το ίδιο αρχικό μη κρυπτογραφημένο σύμβολο, γεγονός που καθιστά δυσκολότερη την κρυπτανάλυση καθώς δεν διατηρούνται οι συχνότητες των γραμμάτων (Λάσκαρη 2010). Τα πολυαλφαβητικά κρυπτογραφήματα είναι οι πιο δυνατοί τύποι κρυπτογραφημάτων αντικατάστασης. Σε τέτοια συστήματα το σχέδιο κειμένου κρυπτογραφείται χρησιμοποιώντας κάθε φορά διαφορετική μονοαλφαβητική αντικατάσταση. Ένα σύνολο κανόνων μονοαλφαβητικής αντικατάστασης δημιουργείται και ένα συγκεκριμένο κλειδί καθορίζει ποιοι κανόνες θα χρησιμοποιούνται κάθε φορά. Ένας από τους πιο απλούς τύπους κρυπτογραφήματος είναι ο Vigenere. Σε αυτό το κρυπτογράφημα το σύνολο των κανόνων αντικατάστασης αποτελείται από 26 Caesar κρυπτογραφήματα, μαζί με μεταβιβάσεις από 0 σε 25. Κάθε κρυπτογράφημα βασίζεται σε ένα κλειδί γράμμα, όπως παρουσιάζεται και στον παρακάτω πίνακα Vigenere. Όπως παρατηρείται από τον ακόλουθο πίνακα, το κάθε γράμμα της αλφαβήτου, ανάλογα με το κλειδί γράμμα που υποδεικνύει την μετακίνηση, αντικαθίσταται στο κρυπτογραφημένο κείμενο με γράμμα που υποδεικνύεται από το κλειδί. Για παράδειγμα, αν υπάρχει μετακίνηση 3 θέσεων, τότε είναι προφανές ότι το γράμμα "a" αντικαθίσταται από το γράμμα "d".

Στον παρακάτω πίνακα 2.1 παρουσιάζεται ο πίνακας Vigenere (Incarnato and Auslander 2003).

Πίνακας 2.1: Ο πίνακας του κρυπτογραφήματος Vigenere.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
B	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
C	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
D	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
E	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
F	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
G	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
H	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
I	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
J	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
K	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
L	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
M	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
N	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
O	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
P	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Q	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
R	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
S	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
T	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
U	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
V	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
W	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
X	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
Y	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
Z	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y

Η αποκρυπτογράφηση είναι εξίσου απλή. Η λέξη κλειδί είναι αυτή που μπορεί να υποδείξει την αντίστοιχη γραμμή του πίνακα με την οποία γίνεται η αποκρυπτογράφηση. Η θέση του γράμματος του κρυπτοκειμένου καθορίζει την στήλη και το γράμμα του σχεδίου του κειμένου είναι στην κορυφή της στήλης.

Σε αντίθεση με τα μονοαλφαβητικά κρυπτογραφήματα, τα πολυαλφαβητικά δεν είναι και τόσο τρωτά στις επιθέσεις συχνοτήτων ανάλυσης. Αλλά το σύστημα μπορεί να σπαστεί σε περίπτωση που ο επιτιθέμενος μαντέψει σωστά το μήκος του κλειδιού. Ένας τρόπος εξάλειψης μιας τέτοιας επίθεσης είναι να χρησιμοποιούνται μη επαναλαμβανόμενα κλειδιά.

2.4.1.2. Κρυπτογραφήματα Μετακίνησης ή Μετάθεσης.

Τα κρυπτογραφήματα μετακίνησης ακολουθούν ειδικές τεχνικές στις οποίες το κανονικό πρότυπο των χαρακτήρων αλλάζει. Ένα απλό παράδειγμα ενός κρυπτογραφήματος μετακίνησης μπορεί να περιλαμβάνει αντιστροφή του μηνύματος ή την διαίρεσή του σε ζευγάρια και ανταλλαγή αυτών. Γενικά οι τεχνικές

που υιοθετούνται είναι: Αντιστροφή μηνύματος, γεωμετρικά πρότυπα, μετακίνηση διαδρομής και κιονοειδής μετακίνηση.
Αφού τα γράμματα του κρυπτοκειμένου είναι τα ίδια σαν αυτά του σχεδίου κειμένου, μια ανάλυση συχνότητας γράμματος καθιστά ικανή επίθεση στο κρυπτογράφημα μετακίνησης. Με σκοπό να εξαλειφθεί μία επιτυχής επίθεση, προτείνεται μία δεύτερη μετάθεση του κρυπτοκειμένου. Γενικά τα κρυπτογραφήματα μετακίνησης είναι τρωτά σε κρυπτανάλυση και μπορεί να απαιτούν μηνύματα με καθορισμένο μήκος. Γι' αυτό τα κρυπτογραφήματα αντικατάστασης είναι πιο πολύ εξαπλωμένα.

2.4.2. Μοντέρνα Κρυπτοσυστήματα.

Στα μοντέρνα Κρυπτοσυστήματα ανήκουν τα Συμμετρικά Κρυπτοσυστήματα και τα Ασύμμετρα κρυπτοσυστήματα.

2.4.2.1. Συμμετρικά Κρυπτοσυστήματα.

Οι συμμετρικοί αλγόριθμοι εμφανίστηκαν στην μηχανοποιημένη μορφή στη δεκαετία του 70. Χωρίζονται σε δύο μεγάλες κατηγορίες ανάλογα με τον τρόπο που χειρίζονται το κείμενο προς επεξεργασία.

- Κρυπταλγόριθμοι ροής (Stream Ciphers).
- Κρυπταλγόριθμοι τμήματος (Block Ciphers).

Οι αλγόριθμοι ροής εμφανίζονται πάνω στην μικρότερη μονάδα ενός ψηφιακού συστήματος τα δυαδικά ψηφία (bits). Αντιθέτως, οι κρυπταλγόριθμοι τμήματος εφαρμόζονται με μονάδα επεξεργασίας τις ψηφιακές λέξεις, δηλαδή συστάδες από δυαδικά ψηφία. Ένας κρυπταλγόριθμος τμήματος είναι μια επαναληπτική διαδικασία μίας κλάσης συναρτήσεων στις οποίες η πληροφορία ρέει διαδοχικά και μετασχηματίζεται. Το αποτέλεσμα αποτελεί την σύνθετη πολλαπλασιαστική δομή κατά Shannon. Η κάθε επανάληψη ονομάζεται γύρος του κρυπταλγορίθμου. Ο καινούριος γύρος τροφοδοτείται με τα αποτελέσματα του προηγούμενου καθώς επίσης και με ένα κλειδί που ονομάζεται κλειδί γύρου. Τα κλειδιά γύρου δημιουργούνται από ένα πρόγραμμα κλειδιού το οποίο συνήθως είναι εκτός του

βασικού αλγορίθμου. Η διαδικασία υπολογισμού των υποκλειδιών κάθε γύρου γίνεται στην αρχή για λόγους ταχύτητας.

Συμμετρικό κρυπτοσύστημα **είναι το σύστημα εκείνο το οποίο χρησιμοποιεί κατά την διαδικασία της κρυπτογράφησης-αποκρυπτογράφησης ένα κοινό κλειδί** (Σχήμα 2.4). Η ασφάλεια αυτών των αλγορίθμων βασίζεται στην μυστικότητα του κλειδιού. Τα συμμετρικά κρυπτοσυστήματα προϋποθέτουν την ανταλλαγή του κλειδιού μέσα από ένα ασφαλές κανάλι επικοινωνίας ή μέσα από την φυσική παρουσία των προσώπων. Αυτό το χαρακτηριστικό καθιστά δύσκολη την επικοινωνία μεταξύ απομακρυσμένων μερών.

Σχήμα 2.5: Το συμμετρικό μοντέλο.

Τα στάδια της επικοινωνίας της παραπάνω εικόνας είναι τα ακόλουθα:

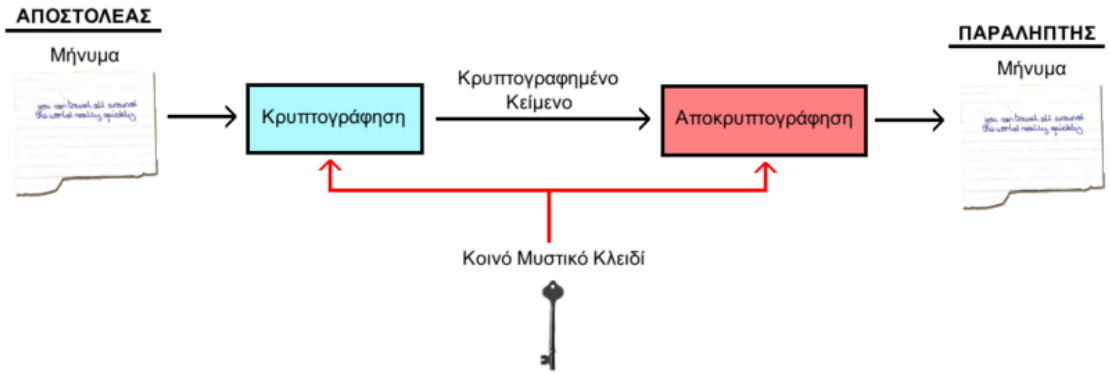
1. Ο αποστολέας ή ο παραλήπτης αποφασίζει για ένα κλειδί το οποίο το επιλέγει τυχαία μέσα από τον κλειδοχώρο.
2. Ο παραλήπτης αποστέλλει το κλειδί στον αποστολέα μέσα από ένα ασφαλές κανάλι.
3. Ο αποστολέας δημιουργεί ένα μήνυμα όπου τα σύμβολα m ανήκουν στον χώρο των μηνυμάτων.
4. Κρυπτογραφεί το μήνυμα με το κλειδί που έλαβε από την παραλήπτη και η παραγόμενη κρυπτοσυμβολοσειρά αποστέλλεται.
5. Ο παραλήπτης λαμβάνει την κρυπτοσυμβολοσειρά και στην συνέχεια με το ίδιο κλειδί την αποκρυπτογραφεί και η έξοδος που παράγεται είναι το μήνυμα.

Καταστάσεις Λειτουργίας Συμμετρικών Κρυπταλγορίθμων

Ο τρόπος που λειτουργούν οι συμμετρικοί κρυπταλγόριθμοι είναι οι τρόποι διασύνδεσης των κρυπταλγόριθμων τμήματος, με στόχο την περαιτέρω αύξηση της κρυπτογραφικής δύναμης και την αποτελεσματικότερη απόκρυψη πιθανών υπολειμμάτων πληροφορίας του απλού κειμένου που μπορεί να υπάρχει στο κρυπτοκείμενο. Υπάρχουν τέσσερις τυποποιημένοι τρόποι λειτουργίας και αρκετοί μη τυποποιημένοι τρόποι λειτουργίας (Κάτος και Στεφανίδης, 2003). Οι τέσσερις τρόποι λειτουργίας είναι:

1. Κατάσταση λειτουργίας ηλεκτρονικού βιβλίου κωδικών (Electronic Codebook, ECB).
2. Κατάσταση λειτουργίας αλυσιδωτής σύνδεσης τμημάτων κρυπτογραφίας (cipher block chaining, CBC).
3. Κατάσταση λειτουργίας κρυπτογραφίας ανάδρασης cipher feedback mode, CFB).
4. Ανάδραση εξόδου output feedback, OFB).

Η κρυπτογράφηση του συμμετρικού κλειδιού (Symmetric Cryptography) βασίζεται στην ύπαρξη ενός και μόνο κλειδιού, το οποίο χρησιμοποιείται τόσο στη κρυπτογράφηση όσο και στην αποκρυπτογράφηση του μηνύματος. Το κλειδί αυτό θα πρέπει να είναι γνωστό μόνο στα συναλλασσόμενα μέρη. Η διαδικασία της κρυπτογράφησης και αποκρυπτογράφησης φαίνεται στο σχήμα που ακολουθεί:



Σχήμα 2.6: Η διαδικασία της κρυπτογράφησης και αποκρυπτογράφησης με την χρήση δημόσιου κλειδιού.

Ένα πρόβλημα το οποίο υφίσταται στους συμμετρικούς αλγορίθμους κρυπτογράφησης είναι η αδυναμία ανταλλαγής του κλειδιού με κάποιο ασφαλή τρόπο. Στην σύγχρονη ψηφιακή εποχή ο αποστολέας και ο παραλήπτης του

μηνύματος πολλές φορές δεν γνωρίζονται, οπότε για την μετάδοση του κλειδιού από τον έναν στον άλλο θα πρέπει να υπάρχει κάποιο ασφαλές κανάλι επικοινωνίας.

Φυσικά το διαδίκτυο δεν μπορεί να αποτελέσει κανάλι ασφαλούς επικοινωνίας, οπότε **η χρήση της συμμετρικής κρυπτογράφησης σε εφαρμογές ηλεκτρονικού εμπορίου, ανταλλαγής ηλεκτρονικών μηνυμάτων κ.τ.λ. ουσιαστικά δεν υφίσταται**.

Το βασικό πλεονέκτημα των αλγορίθμων συμμετρικού κλειδιού είναι ότι η διαδικασία της κρυπτογράφησης και αποκρυπτογράφησης είναι πολύ γρήγορη και δεν καταναλώνει σημαντική υπολογιστική ισχύ.

Οι πιο γνωστοί αλγόριθμοι αυτού του είδους είναι οι DES, Tripple DES, IDEA, RC2, RC4, AES.

Παράδειγμα από την καθημερινή ζωή:

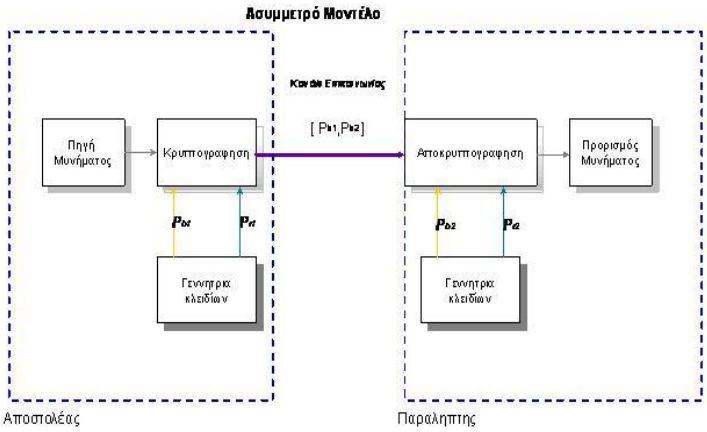
Θα παρουσιάσουμε ένα αναλογικό παράδειγμα από την καθημερινή ζωή το οποίο περιγράφει την κρυπτογράφηση συμμετρικού κλειδιού. Έστω ο χρήστης_1 και ο χρήστης_2, θέλουν να επικοινωνήσουν με ασφάλεια χρησιμοποιώντας το δημόσιο ταχυδρομείο. Ο χρήστης_1 θέλει να στείλει ένα καμουφλαρισμένο-κρυφό μήνυμα στον χρήστη_2 και περιμένει μια καμουφλαρισμένη-κρυφή απάντηση από αυτόν.

Σύμφωνα με την κρυπτογράφηση συμμετρικού κλειδιού ο χρήστης_1 θα βάλει το μήνυμά του μέσα σε ένα κουτί με λουκέτο για το οποίο έχει το κλειδί. Στέλνει το κλειδωμένο κουτί με το δημόσιο ταχυδρομείο στον χρήστη_2. Ο χρήστης_2 έχει ένα ίδιο κλειδί (το οποίο έχει πάρει από τον χρήστη_1 στο παρελθόν, σε διαπροσωπική συνάντηση) και μόλις λαμβάνει το κουτί, το ανοίγει με τη βοήθεια του κλειδιού και διαβάζει το μήνυμα. Ο χρήστης_2 βάζει το μήνυμά του στο κουτί, το κλειδώνει και το στέλνει με το δημόσιο ταχυδρομείο στον χρήστη_1.

Το πρόβλημα που παρουσιάζεται είναι ότι ενώ το κλειδί είναι κοινό και για τους δύο, για να δώσει κάποιος αντίγραφο του κλειδιού στον άλλο θα πρέπει να συναντηθούν γιατί δεν είναι ασφαλές να το στείλουν με το δημόσιο ταχυδρομείο, καθώς κάποιος άλλος χρήστης θα μπορούσε να το υποκλέψει. Μια τέτοια ενέργεια θα σήμαινε ότι ο τρίτος χρήστης θα μπορούσε να δημιουργήσει αντίγραφο ώστε στο μέλλον να υποκλέπτει ή να παραποιεί τα μηνύματα που ανταλλάσσονται μεταξύ των δύο χρηστών.

2.4.2.2. Ασύμμετρα Κρυπτοσυστήματα.

Το ασύμμετρο κρυπτοσύστημα ή κρυπτοσύστημα δημοσίου κλειδιού δημιουργήθηκε για να καλύψει την αδυναμία μεταφοράς κλειδιών που παρουσίαζαν τα συμμετρικά συστήματα. Χαρακτηριστικό του είναι ότι έχει δυο είδη κλειδιών ένα ιδιωτικό και ένα δημόσιο. Το δημόσιο είναι διαθέσιμο σε όλους ενώ το ιδιωτικό είναι μυστικό. Η βασική σχέση μεταξύ τους είναι : ό,τι κρυπτογραφεί το ένα, μπορεί να το αποκρυπτογραφήσει μόνο το άλλο.



Εικόνα 2.6: Μετάδοση μηνύματος με χρήση ιδιωτικού κλειδιού.

Τα στάδια της επικοινωνίας της παραπάνω εικόνας είναι τα ακόλουθα:

- Η γεννήτρια κλειδιών του αποστολέα παράγει 2 ζεύγη κλειδιών.
- Η γεννήτρια κλειδιών του παραλήπτη παράγει 2 ζεύγη κλειδιών.
- Ο αποστολέας και ο παραλήπτης ανταλλάσσουν τα δημόσια ζεύγη.
- Ο αποστολέας δημιουργεί ένα μήνυμα όπου τα σύμβολα m ανήκουν στον χώρο των μηνυμάτων.
- Κρυπτογραφεί το μήνυμα με το δημόσιο κλειδί του παραλήπτη και η παραγόμενη κρυπτοσυμβολοσειρά αποστέλλεται.
- Ο παραλήπτης λαμβάνει την κρυπτοσυμβολοσειρά και στη συνέχεια με το ιδιωτικό της κλειδί την αποκρυπτογραφεί και η έξοδος που παράγεται είναι το μήνυμα.

2.5. Ο αλγόριθμος DES

2.5.1. Ο αλγόριθμος DES ως πρότυπο.

Το Data Encryption Standard (DES), αποτέλεσε και αποτελεί ακόμη και σήμερα το πιο δημοφιλές και διαδεδομένο σύστημα κρυπτογράφησης το οποίο υιοθετήθηκε το 1977 από το National Bureau of Standards (NBS) των ΗΠΑ, και έγινε αργότερα γνωστό ως National Institute of Standards and Technology (NIST), ως πρότυπο επεξεργασίας πληροφοριών των ομοσπονδιακών αρχών (Federal Information Processing Standard 46-FIPS PUB 46) (Jorstad and Landgrave 1997) και για γενική χρήση.

Πιο συγκεκριμένα, παρά τις επικρίσεις που δέχθηκε αυτός ο αλγόριθμος, ο DES εγκρίθηκε ως ομοσπονδιακό πρότυπο τον Νοέμβριο του 1976 και δημοσιεύθηκε στις 15 Ιανουαρίου του 1977 ως FIPS PUB 46 και η χρήση του ήταν επιτρεπτή σε όλα τα μη απόρρητα δεδομένα. Στη συνέχεια επιβεβαιώθηκε ως πρότυπο το 1983, το 1988 (αναθεωρήθηκε ως FIPS-46-1), το 1993 (ως FIPS-46-2) και πάλι το 1999 (ως FIPS-46-3). Ο τελευταίος ορισμός ήταν ο Triple DES. Στις 26 Μαΐου του 2002 ο DES τελικά εκτοπίστηκε από τον Advanced Encryption Standard (AES) κατόπιν δημόσιου διαγωνισμού. Στις 19 Μαΐου του 2005 ο FIPS 46-3 είχε επισήμως αποσυρθεί, αλλά το Εθνικό Ινστιτούτο Προτύπων και Τεχνολογίας (National Institute of Standards and Technology - NIST) ενέκρινε τον Triple DES στο έτος 2003 για τις ευαίσθητες πληροφορίες της κυβέρνησης. Μια άλλη θεωρητική επίθεση, η γραμμική κρυπτανάλυση, δημοσιεύθηκε το 1994, αλλά ήταν μια επίθεση «brute force» το 1998 που αναπαράστησε/απέδειξε ότι μπορεί κάποιος θα μπορούσε πρακτικά να επιτεθεί στον DES και τονίστηκε η ανάγκη για αντικατάσταση του αλγόριθμου (Veurink, et. al., 2005).

Η δημιουργία του DES θεωρείται καταλυτικής σημασίας για την ακαδημαϊκή μελέτη του συστήματος κρυπτογραφίας, ιδιαίτερα όσον αφορά τις μεθόδους αποκρυπτογράφησης των κρυπτογραφικών block.

Μπορεί να ειπωθεί ότι το "αρχικό άλμα" του DES ξεπέρασε τις στρατιωτικές μελέτες και την ανάπτυξη των αλγορίθμων κρυπτογράφησης. Στην δεκαετία του 1970 υπήρχαν πολύ λίγοι κρυπτογράφοι, εκτός εκείνων των στρατιωτικών ή των μυστικών οργανώσεων, και ελάχιστη ήταν η ακαδημαϊκή έρευνα της κρυπτογραφίας. Σήμερα υπάρχουν πολλοί δραστήριοι ακαδημαϊκοί κρυπτολόγοι

και τμήματα μαθηματικών με ισχυρά προγράμματα στην κρυπτογραφία και την ασφάλεια των πληροφοριών και των εμπορικών εταιρειών και συμβούλων. Μια γενιά κρυπταναλυτών έχει αναλύσει εξονυχιστικά τον αλγόριθμο DES προσπαθώντας να τον "σπάσουν". Ανέφεραν πως ο DES έκανε περισσότερα για να προάγει τον τομέα της κρυπτανάλυσης από οτιδήποτε άλλο γιατί υπήρχε ένας αλγόριθμος για μελέτη. Ένα εκπληκτικό μερίδιο της βιβλιογραφίας στην κρυπτογραφία κατά τη δεκαετία του 1970 και του 1980 ασχολήθηκε με τον DES και ο DES είναι πρότυπο για σύγκριση για όλους τους αλγόριθμους συμμετρικού κλειδιού.

2.5.2. Βελτιώσεις του DES

2.5.2.1. Advanced Encryption Standard (AES).

Στην κρυπτογραφία, το **Advanced Encryption Standard (AES)** είναι ένα πρότυπο συμμετρικού κλειδιού κρυπτογράφησης που έχει εγκριθεί από την κυβέρνηση των ΗΠΑ. Το πρότυπο αυτό περιλαμβάνει τρεις κρυπταλγόριθμους , συγκεκριμένα τα AES-128, AES-192 και AES-256, που εγκρίθηκε από μια μεγαλύτερη συλλογή που δημοσιεύθηκε αρχικά ως **Rijndael.** Κάθε ένας από αυτούς τους αλγορίθμους έχει μέγεθος μπλοκ-128, με βασικά μεγέθη των 128, 192 και 256 bits, αντίστοιχα. Οι AES κρυπταλγόριθμοι έχουν αναλυθεί εκτενώς και σήμερα χρησιμοποιούνται παγκοσμίως, όπως και ο προκάτοχός τους, ο αλγόριθμος DES.

Το AES ανακοινώθηκε από το Εθνικό Ινστιτούτο Προτύπων και Τεχνολογίας (National Institute of Standards and Technology-NIST) των ΗΠΑ ως FIPS PUB 197 (FIPS 197) στις 26 Νοεμβρίου 2001 μετά από μια 5-ετή διαδικασία τυποποίησης στην οποία συμμετείχαν δεκαπέντε ανταγωνιστικά σχέδια τα οποία παρουσιάστηκαν και αξιολογήθηκαν πριν το Rijndael να το επιλέξει ως το πιο κατάλληλο. Έγινε πρότυπο από την ομοσπονδιακή κυβέρνηση στις 26 του Μαΐου 2002 μετά την έγκριση από την Γενική Γραμματέα Εμπορίου. Είναι διαθέσιμο σε πολλά διαφορετικά πακέτα κρυπτογράφησης. Το AES είναι ο πρώτος αλγόριθμος που έχει δημόσια και ανοικτή πρόσβαση όπως εγκρίθηκε από την NSA για άκρως απόρρητες πληροφορίες.

Περιγραφή κρυπτογράφησης του AES

Το πρότυπο AES περιγράφει μια συμμετρική διαδικασία κρυπτογράφησης μυστικού κλειδιού. Το πρότυπο υποστηρίζει την χρήση κλειδιών μήκους 128,192, και 256 bits. Ανάλογα με το ποιο μήκος κλειδιού χρησιμοποιείται, συνήθως χρησιμοποιείται η συντόμευση AES-128, AES-192 και AES-256 αντίστοιχα. Ανεξάρτητα από το μήκος κλειδιού, ο αλγόριθμος επενεργεί πάνω σε μπλοκ δεδομένων μήκους 128 bits. Η διαδικασία κρυπτογράφησης είναι επαναληπτική. Αυτό σημαίνει ότι σε κάθε μπλοκ δεδομένων γίνεται μια επεξεργασία η οποία επαναλαμβάνεται από έναν αριθμό από φορές, ανάλογα με το μήκος του κλειδιού. Κάθε επανάληψη ονομάζεται γύρος (round). Στον πρώτο γύρο επεξεργασίας ως είσοδος είναι ένα κείμενο (plaintext) και το αρχικό κλειδί, ενώ στους γύρους που ακολουθούν ως είσοδος είναι το μπλοκ που έχει προκύψει από τον προηγούμενο γύρο καθώς και ένα κλειδί που έχει παραχθεί από το αρχικό με βάση κάποια διαδικασία που ορίζει ο αλγόριθμος. Το τελικό προϊόν της επεξεργασίας είναι το κρυπτογραφημένο μπλοκ (ciphertext). Το μπλοκ αυτό πρέπει να σημειωθεί ότι έχει ακριβώς το ίδιο μέγεθος (128 bits) με το αρχικό κείμενο.
Όπως αναφέρθηκε παραπάνω, ο AES τροφοδοτείται με ακολουθίες από bits των 128 bits (μπλοκ) καθώς και από κλειδιά, που μπορεί να έχουν μέγεθος 128,192 ή 256 bits. Τα κλειδιά αυτά ονομάζονται κλειδιά κρυπτογράφησης (cipher keys) για να διαχωριστούν από τα κλειδιά που παράγονται κατά την λειτουργία του αλγορίθμου.
Όλες οι λειτουργίες που επιτελεί ο αλγόριθμος αυτός γίνονται πάνω σε ένα δισδιάστατο πίνακα που αποκαλείται Κατάσταση (State). Ο πίνακας αυτός περιλαμβάνει τέσσερεις γραμμές από bytes, με κάθε μία γραμμή να αποτελείται από **Nb** Bytes. Ο αριθμός που αντιστοιχεί στην ποσότητα Nb υπολογίζεται αν διαιρεθεί το μήκος τους μπλοκ με το 32. Εφόσον στον AES υποστηρίζονται μπλοκ μεγέθους μόνο 128 bits, το Nb θα έχει τιμή 4.

2.5.2.2. Triple DES.

Ο αλγόριθμος Triple DES (γνωστός ως 3DES, 3-DES, TDES) βασίζεται στον αλγόριθμο DES. Έναντι τοθ αλγορίθμου DES, ο Triple DES έχει το πλεονέκτημα

της μεγαλύτερης αξιοπιστίας από την χρήση μεγαλύτερου κλειδιού, γεγονός που αποτρέπει περισσότερες επιθέσεις στο κρυπτοκείμενο. Ωστόσο, ακόμη και αυτή η βελτιωμένη έκδοση του αλγορίθμου DES δεν είναι αρκετά ισχυρή για να προστατεύσει τα δεδομένα για μεγάλο χρονικό διάστημα. Ο αλγόριθμος DES θεωρείται παλιός και πρέπει να αντικατασταθεί. Έτσι λοιπόν, το Ινστιτούτο Προτύπων και Τεχνολογίας (NIST) έχει ορίσει τον αλγόριθμο AES ως αντικαταστάτη του DES. Συνεπώς ο αλγόριθμος Triple DES έχει οριστεί από τον NIST ως μία προσωρινή λύση έως ότου τελειοποιηθεί ο αλγόριθμος AES. Ο αλγόριθμος AES είναι πιο γρήγορος και πιο ισχυρός έναντι του Triple DES.

Ο Triple DES είναι μία διαφορετική χρήση του DES καθώς χειρίζεται 3 κλειδιά των 64-bit για ένα σύνολο κλειδιού 192-bit. Ο Triple DES εκτελείται τρείς φορές πιο αργά από τον DES, αλλά είναι πιο ασφαλής αν χρησιμοποιηθεί σωστά.

Η διαδικασία της κρυπτογράφησης είναι ίδια με αυτή που πραγματοποιείται από τον αλγόριθμο DES, με την διαφορά ότι η διαδικασία αυτή εκτελείται τρείς φορές, εξ ου και το όνομα «Triple DES». Η διαδικασία της αποκρυπτογράφησης είναι παρόμοια με την διαδικασία της κρυπτογράφησης με την διαφορά ότι εκτελείται αντίστροφα.

2.6. Σύγκριση συμμετρικής και ασύμμετρης κρυπτογράφησης.

Η συμμετρική κρυπτογράφηση κλειδιού έχει ένα μειονέκτημα. Δύο άνθρωποι που επιθυμούν να ανταλλάξουν εμπιστευτικά μηνύματα πρέπει να μοιραστούν ένα μυστικό κλειδί. Το κλειδί πρέπει να ανταλλαχθεί με έναν ασφαλή τρόπο και όχι με τα μέσα που θα επικοινωνούσαν κανονικά. Αυτό συνήθως είναι το πιο δύσκολο σημείο και το σύστημα κρυπτογραφίας δημόσιων κλειδιών (ή ασύμμετρο) παρέχει μια εναλλακτική λύση.

Γενικά, οι τεχνικές δημόσιου κλειδιού είναι πολύ ισχυρότερες υπολογιστικά από τους καθαρά συμμετρικούς αλγόριθμους, αλλά η σωστή χρήση αυτών των τεχνικών επιτρέπει μια ευρεία ποικιλία εφαρμογών τους.

Όσον αφορά την ασφάλεια, δεν υπάρχει τίποτα που να καθιστά ασφαλέστερους τους αλγορίθμους δημόσιου κλειδιού σε σύγκριση με τους συμμετρικούς αλγορίθμους κλειδιών. Υπάρχουν δημοφιλείς και μη δημοφιλείς αλγόριθμοι. Υπάρχουν αυτοί που έχουν παραβιαστεί και αυτοί που δεν έχουν ακόμα τουλάχιστον σπάσει. Δυστυχώς, η δημοτικότητα δεν είναι ένας αξιόπιστος δείκτης

της ασφάλειας. Θα μπορούσε να θεωρηθεί ότι το «σπάσιμο» ενός αλγορίθμου, όσον αφορά κάποιους καθορισμένους με σαφήνεια στόχους ασφαλείας, είναι ισοδύναμο με την επίλυση ενός από τα δημοφιλέστερα μαθηματικά προβλήματα που θεωρούνται ότι είναι ισοδύναμα με το πρόβλημα του διακριτού λογαρίθμου. Γενικά, κανένας από αυτούς τους αλγορίθμους δεν έχει αποδειχθεί να είναι απόλυτα ασφαλής..

2.7. ΕΠΙΛΟΓΟΣ

Σε αυτό το κεφάλαιο έγινε αναφορά στην επιστήμη της κρυπτολογίας, η οποία διαχωρίζεται στις κατηγορίες της κρυπτογραφίας και της κρυπτανάλυσης. Δόθηκαν οι απαραίτητοι ορισμοί που χαρακτηρίζουν τις δύο αυτές κατηγορίες και παρουσιάστηκαν τα είδη κρυπτοσυστημάτων. Επίσης έγινε αναφορά στην σπουδαιότητα και την ανάγκη για την ύπαρξη της κρυπτολογίας. Ακόμη έγινε αναφορά στους σημαντικότερους αλγορίθμους κρυπτογράφησης οι οποίοι είναι οι DES, AES, και ο Triple DES.

Όπως εξετάστηκε, η χρήση της κρυπτογραφίας είναι σημαντική για την ασφάλεια των δεδομένων, ωστόσο κατά την διαδικασία αυτής η δομή ενός κειμένου μεταβάλλεται.

Ο βαθμός της επιρροής της κρυπτογραφίας στη συχνότητα και κατανομή των γραμμάτων θα εξεταστεί στο επόμενο κεφάλαιο με τη βοήθεια λογισμικού.

ΚΕΦΑΛΑΙΟ 3

3. ΕΦΑΡΜΟΓΕΣ ΛΟΓΙΣΜΙΚΟΥ

3.1. ΕΙΣΑΓΩΓΗ

Σε αυτό το κεφάλαιο θα παρουσιαστούν και θα αναλυθούν οι εφαρμογές που αναπτύχθηκαν στα πλαίσια αυτής της πτυχιακής. Συγκεκριμένα θα παρουσιαστούν δύο εφαρμογές οι οποίες συνδυάζουν τον τομέα της κρυπτογραφίας με την συχνότητα και κατανομή γραμμάτων, δηλαδή κατά πόσο επηρεάζει η κρυπτογραφία την κατανομή γραμμάτων. Η πρώτη εφαρμογή έχει ως στόχο να αναγνωρίσει την αρχική γλώσσα προέλευσης ενός κρυπτογραφημένου κειμένου, και η δεύτερη εξετάζει την επιρροή στην συχνότητα και κατανομή γραμμάτων έπειτα από το πέρασμά τους από τον αλγόριθμο κρυπτογράφησης AES. Μέσω αυτών των εφαρμογών μπορεί να υποδειχτεί η γλώσσα προέλευσης του αρχικού κειμένου, όχι απόλυτα αλλά κατά ένα μεγάλο ποσοστό, από την εξαγωγή ποσοστών συχνότητας των γραμμάτων του αρχικού κειμένου. Εν συνεχεία θα γίνει αναφορά των αποτελεσμάτων που εξάγονται πειραματικά, ώστε να προβληθούν χρήσιμα συμπεράσματα.

3.2. Πρόλογος

Οι παρακάτω εφαρμογές στις οποίες θα αναφερθούμε είναι γραμμένες σε γλώσσα **Java** και έχουν αναπτυχθεί στο περιβάλλον του **NetBeans 6.9.1**. Η επιλογή της γλώσσας προγραμματισμού Java έγινε καθώς θεωρήθηκε προσιτή και εύχρηστη. Το προγραμματιστικό περιβάλλον του NetBeans επιλέχθηκε καθώς παρέχει αρκετές ευκολίες και υποδείξεις κατά την διαδικασία ανάπτυξης του λογισμικού σε γλώσσα Java. Αυτό είναι γενικό παραδεκτό από προγραμματιστές λογισμικού καθώς παρέχει διάφορες δυνατότητες όπως επεξήγηση των λαθών κατά την εισαγωγή μεθόδων και αντικειμένων στις κλάσεις που δημιουργούνται, αλλά και προτροπής εισαγωγής πακέτων για την υποστήριξη κάποιων μεθόδων.

3.3. Αναγνώριση προέλευσης του αρχικού κειμένου από ένα κρυπτοκείμενο.

Όπως αναφέρθηκε παραπάνω ένας από τους στόχους είναι να δημιουργηθεί μια εφαρμογή η οποία θα εξετάζει ένα κρυπτοκείμενο σχετικά με την γλώσσα προέλευσής του. Συγκεκριμένα, στο προγραμματιστικό περιβάλλον του NetBeans, αναπτύχθηκε μία εφαρμογή η οποία δέχεται ως είσοδο ένα απλό κείμενο γραμμένο στην ελληνική ή την αγγλική γλώσσα. Το κείμενο αυτό κρυπτογραφείται και μετατρέπεται σε μία μορφή η οποία είναι αδύνατο να υποδείξει την αρχική γλώσσα προέλευσής του. Σημειώνεται ότι η γραφή του κρυπτοκειμένου αποτελείται από διάφορα σύμβολα, μιας πληθώρας από αυτά που είναι διαθέσιμα από το προγραμματιστικό περιβάλλον του NetBeans. Εν συνεχεία, αυτό που εξετάζεται από την εφαρμογή είναι το κρυπτοκείμενο το οποίο αποθηκεύεται σε ένα αρχείο με το όνομα "Encrypted". Το λογισμικό που αναπτύχθηκε κάνει αναζήτηση μόνο στο κρυπτοκείμενο, έχοντας πλήρη άγνοια για το αρχικό κείμενο που εισήχθη στο αρχείο εισόδου "Caesar.txt". Αυτό μπορεί να προέρχεται από οποιαδήποτε γλώσσα και αυτό που αναζητούμε είναι η αρχική γλώσσα προέλευσης. Η γλώσσα δηλαδή στην οποία είναι γραμμένο το αρχικό κείμενο (plain text, όπως έχει αναφερθεί στο κεφάλαιο 2). Φυσικά το να δημιουργηθεί μια εφαρμογή η οποία μπορεί να περιλαμβάνει όλες τις γλώσσες και να ανιχνεύει μέσω του κρυπτοκειμένου την γλώσσα προέλευσης του αρχικού κειμένου είναι ιδιαίτερα δύσκολο, καθώς θα πρέπει να περιληφθεί μεγάλη βάση δεδομένων για όλα τα γράμματα κάθε γλώσσας. ***Έτσι λοιπόν η εφαρμογή που αναπτύχθηκε είναι σε θέση να ανιχνεύσει την αρχική γλώσσα προέλευσης του κρυπτοκειμένου, δεδομένου όμως ότι αυτό θα προέρχεται από την αγγλική ή την ελληνική γλώσσα***.

3.3.1. Επεξήγηση του κώδικα αναγνώρισης της αρχικής γλώσσας προέλευσης κρυπτοκειμένου.

Δημιουργείται μία κλάση με όνομα "CaesarCypher" και ορίζονται δύο μεταβλητές file και file2 στις οποίες δηλώνονται τα αρχεία "Ceasar.txt" και "Encrypted.doc" αντίστοιχα με τις εντολές:

```
static File file = new File("Ceasar.txt");
```

```
static File file2 = new File("Encrypted.doc");
```

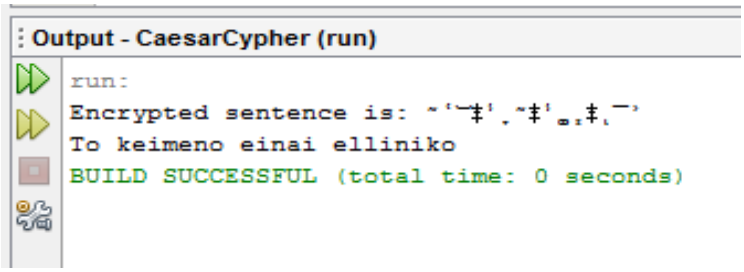
Το "Ceasar.txt" θα χρησιμοποιηθεί για την ανάγνωση ενός κειμένου και το "Encrypted.doc" για την αποθήκευση του κρυπτοκειμένου που θα προκύπτει. Επίσης δηλώνεται και η random τιμή "generator" με την εντολή

```
static Random generator = new Random();
```

η οποία χρησιμοποιείται στην κλάσης της main για να ελέγξει τις τιμές r και r2 στις οποίες θα δίνεται τυχαία ένας αριθμός από ένα εύρος 33 χαρακτήρων. Το εύρος αυτό έχει επιλεχθεί σύμφωνα με τον πίνακα ASCII (βλ. Παράρτημα Α'). Έτσι λοιπόν, όταν ξεκινάει το σάρωμα του κειμένου στην main, ξεκινάει με βήμα 1 και φτάνει έως το -2 του συνολικού μήκους των χαρακτήρων του κειμένου,

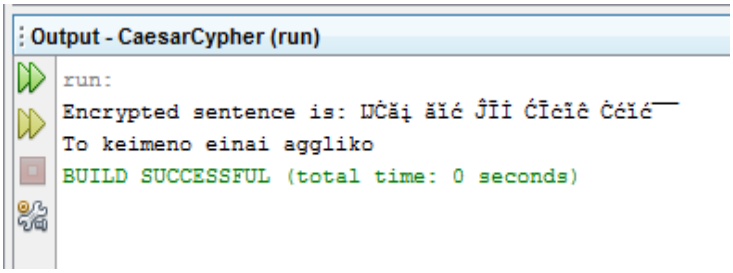
```
for (int iteration = 1; iteration < plainText.length()-2;
iteration++) {
```

ώστε να αποφευχθεί η εκτύπωση επιπρόσθετων συμβόλων στο κρυπτοκείμενο, τα οποία δεν κρυπτογραφούν τίποτα. *Αυτό έχει διαπιστωθεί από πειραματισμό κατά την διαδικασία της κρυπτογράφησης, με έλεγχο μικρού μήκους χαρακτήρων.* Με βάση τον πίνακα ASCII, οι αγγλικοί χαρακτήρες βρίσκονται κάτω από την θέση 122. Γίνεται επίσης έλεγχος, αν ο χαρακτήρας βρίσκεται μεταξύ των θέσεων 85 και 105 του πίνακα ASCII, τότε να γίνει μετατόπιση κατά 130 θέσεις, έτσι ώστε να αποφευχθεί η μετατροπή ενός κεφαλαίου γράμματος να αντικατασταθεί από ένα μικρό γράμμα της ίδιας γλώσσας, γιατί τότε στην ουσία θα γίνουν μετατοπίσεις των γραμμάτων της ίδιας γλώσσας και θα είναι προφανής η γλώσσα προέλευσης του κρυπτοκειμένου. Επίσης γίνεται έλεγχος, αν ο χαρακτήρας βρίσκεται μεταξύ των θέσεων 107 με 122, τότε θα γίνει μετατόπιση κατά 155 θέσεις προς τα πάνω. Έχοντας ελέγξει τις θέσεις που βρίσκονται τα ελληνικά και τα αγγλικά γράμματα και έχοντας ορίσει συγκεκριμένες μετατοπίσεις, γίνεται ο τελικός έλεγχος αν ο χαρακτήρας του κρυπτοκειμένου βρίσκεται σε θέση μεγαλύτερη του 322. Σε αυτή την περίπτωση σημαίνει πως το κρυπτοκείμενο προέρχεται από την ελληνική γλώσσα, όπως φαίνεται και στην εικόνα 3.1 παρακάτω από την εκτέλεση της εφαρμογής,



Εικόνα 3.1: Αναγνώριση ελληνικού κρυπτογραφήματος.

διαφορετικά προέρχεται από την αγγλική, εικόνα 3.2.



Εικόνα 3.2: Αναγνώριση αγγλικού κρυπτογραφήματος.

Αναλυτικά ο κώδικας της εφαρμογής αναγνώρισης της γλώσσας προέλευσης ενός κρυπτοκειμένου βρίσκεται στο παράρτημα Β'.

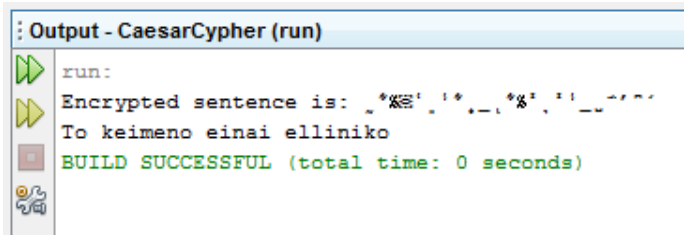
Επίσης πρέπει να τονίζουμε ότι μέσα από αυτή την εφαρμογή εκτελείται εν μέρει η διαδικασία της κρυπτογράφησης. Να διευκρινίσουμε ότι λέμε πως η κρυπτογραφία εκτελείται «εν μέρει», διότι χρησιμοποιείται ένα αρχικό κείμενο και αυτό κρυπτογραφείται. Σε αυτή τη φάση, κρατάμε αυτό το μέρος μόνο της κρυπτογραφίας και ασχολούμαστε με την αναγνώριση της γλώσσας του αρχικού κειμένου. **Επισημαίνεται ότι η αναγνώριση της γλώσσας αποτελεί εφαρμογή που δημιουργήθηκε στα πλαίσια της πτυχιακής εργασίας και δεν ανήκει στην διαδικασία της κρυπτογραφίας.** Για να θεωρηθεί ολοκληρωμένη η διαδικασία της κρυπτογραφίας θα πρέπει το κρυπτογραφημένο κείμενο να αποκρυπτογραφείται και να παράγεται σε ένα νέο αρχείο το αποκρυπτογραφημένο κείμενο, το οποίο φυσικά θα πρέπει να ταυτίζεται με το αρχικό. Συγκεκριμένα, στη εφαρμογή αυτή, εισάγεται ένα απλό κείμενο στο αρχείο "Caesar.txt", είτε ελληνικό είτε αγγλικό. Εδώ πρέπει να διευκρινιστεί ότι το αρχείο δημιουργείται από την εφαρμογή, ωστόσο η εισαγωγή δεδομένων πρέπει να γίνει από τον χρήστη της εφαρμογής. Εν συνεχεία το κείμενο του "Caesar.txt" κρυπτογραφείται και σε κάθε εκτέλεση της εφαρμογής, στο ίδιο κείμενο κάθε φορά που παράγεται και διαφορετικό κρυπτοκείμενο. Παρακάτω παρατίθεται ένα παράδειγμα από την εκτέλεση της εφαρμογής σε ένα κείμενο λίγων χαρακτήρων.

Στον πίνακα παρακάτω υπάρχει μια φράση (The congress is postponed) η οποία κρυπτογραφείται με διαφορετικούς χαρακτήρες έπειτα από κάθε εκτέλεση.

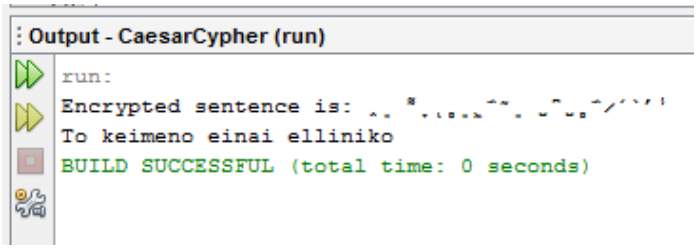
Πίνακας 3.1: Διαφορετικές κρυπτογραφήσεις.

Εκτέλεση	Κρυπτογράφηση
1η	¼Ăÿ□ýĢġāĥÿĦĦ□ĂûĦ□üÿÿġ□ģĢĦħģĢġÿþ
2η	¹ÿü□úğĞþĢüģģ□ÿøģ□ùüüĞ□ĠğģĤĠğĞüû
3η	»āþ□üġĠĀĤþĥĥ□āúĥ□ûþþĠ□ĢġĥĦĢġĠþý

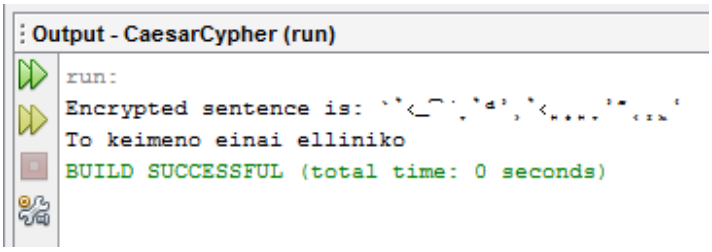
Ομοίως η έκφραση στα ελληνικά (δηλ, Το συμβούλιο αναβλήθηκε) παράγει διαφορετικά σύμβολα τα οποία συνθέτουν το κρυπτογραφημένο κείμενο, όπως παρουσιάζονται ενδεικτικά στις παρακάτω εικόνες.



Εικόνα 3.3: Κρυπτοκείμενο μετά από την πρώτη εκτέλεση.



Εικόνα 3.4: Κρυπτοκείμενο μετά από την δεύτερη εκτέλεση.



Εικόνα 3.5: Κρυπτοκείμενο μετά από την τρίτη εκτέλεση.

Θεωρείται σκόπιμο να τονιστεί ότι οι χαρακτήρες και τα σύμβολα που παράγονται από το NetBeans δεν μεταφέρονται σε έγγραφο του word ή παρουσιάζονται αλλοιωμένα και δεν αναπαριστούν το ακριβές κρυπτοκείμενο. Για τον λόγο αυτό έγινε χρήση εικόνων για την παρουσίασή τους. Επίσης πρέπει να επισημανθεί ότι στα δεδομένα που εισήχθησαν δεν έγινε καμία χρήση συμβόλου (. , ! κτλ) στο τέλος της πρότασης, καθώς αυτό αποτελεί έλεγχο από το λογισμικό για την αναγνώριση της γλώσσας του κρυπτοκειμένου και συνεπώς θα άλλαζε την προέλευσή του. Εμπειρικά παρατηρήθηκε ότι αν γίνει χρήση συμβόλου στο τέλος της πρότασης τότε άσχετα με το συνολικό κείμενο, αυτό θα αναγνωρίζεται ως αγγλικό. Έτσι λοιπόν, για να εξασφάλιση της εγγύησης του αποτελέσματος που προκύπτει, δεν θα πρέπει να γίνεται χρήση συμβόλου στο τέλος της πρότασης στο αρχείο εισαγωγής δεδομένων, το "Caesar.txt".
Η συνολική διαδικασία της κρυπτογράφησης είναι σαφής και ολοκληρωμένη στην επόμενη εφαρμογή, όπου επιπροσθέτως μετρώνται και οι χαρακτήρες του κειμένου πριν και μετά την επεξεργασία της κρυπτογράφησης.

3.4. Επιρροή της συχνότητας και κατανομής γραμμάτων μετά από το πέρασμα ενός αλγορίθμου.

Όπως είχε αναφερθεί ο κύριος στόχος της μελέτης έχει να κάνει με τον συνδυασμό της κρυπτογραφίας με την συχνότητα και την κατανομή των γραμμάτων. Συγκεκριμένα, όπως θα αναλυθεί στην δεύτερη εφαρμογή λογισμικού που αναπτύχθηκε, θα παρουσιαστεί αν και πόσο επηρεάζει η κρυπτογραφία την συχνότητα και την κατανομή των γραμμάτων τόσο στην αγγλική όσο και στην ελληνική γλώσσα. Αναφέρεται επίσης ότι η δεύτερη έχει υλοποιηθεί στο προγραμματιστικό περιβάλλον NetBeans 6.9.1. όμως λόγω κάποιων διαδικαστικών προβλημάτων που αντιμετωπίστηκαν κατά την εκτέλεση, όπως θα αναλυθούν παρακάτω, η εφαρμογή υλοποιείται από δύο προγράμματα.

3.4.1. Περιγραφή της λειτουργίας των κωδικών μέτρησης γραμμάτων κατά την κρυπτογράφηση.

Η βασική ιδέα είναι βασισμένοι στον πηγαίο κώδικα του κρυπταλγόριθμου AES, να προστεθούν στοιχεία έτσι ώστε να εκτελείται η διαδικασία της κρυπτογράφησης και

αποκρυπτογράφησης, αλλά ταυτόχρονα να μπορεί να ελεγχτεί ο αριθμός εμφάνισης των γραμμάτων, ως χαρακτήρες ξεχωριστά πριν και μετά την επεξεργασία της κρυπτογράφησης. Επί της ουσίας δημιουργείται το αρχείο "DESTest.txt" το οποίο λειτουργεί ως αρχείο εισόδου προς επεξεργασία. Σε αυτό το αρχείο μπορούν να εισαχθούν κείμενα μεγάλου μήκους χαρακτήρων είτε ελληνικά, είτε αγγλικά. Με την διαδικασία της κρυπτογράφησης, το αρχικό κείμενο που θα εισαχθεί στο "DESTest.txt" θα κρυπτογραφηθεί, θα μετατραπεί δηλαδή σε διάφορα σύμβολα, και θα αποθηκευτεί στο αρχείο "Encrypted.txt". Για να θεωρηθεί ολοκληρωμένη η διαδικασία της κρυπτογράφησης, θα πρέπει το κρυπτοκείμενο που βρίσκεται στο "Encrypted.txt" να αποκρυπτογραφηθεί και να επέλθει το αρχικό κείμενο, όπως εισήχθη από τον χρήστη. Το αρχικό κείμενο που έχει αποκρυπτογραφηθεί μέσω του AES αποθηκεύεται στο αρχείο "Decrypted.txt". Αυτή η διαδικασία αποτελεί ολοκληρωμένη την διαδικασία της κρυπτογράφησης και αποκρυπτογράφησης, αλλά αυτό το οποίο επιπρόσθετα εξετάζεται προς μελέτη είναι ο ακριβής αριθμών των γραμμάτων που εισήχθησαν και αν αυτός, μεταβάλλεται ή παραμένει ο ίδιος κατά την διαδικασία της κρυπτογράφησης. Συγκεκριμένα, απαιτείται μία σύγκριση μεταξύ του αριθμό των χαρακτήρων του "DESTest.txt" και "Encrypted.txt". Δηλαδή ενώ δίνεται ένα κείμενο εισόδου ζητείται να εξεταστεί αν μέσω της κρυπτογραφίας μεταβάλλεται ο αριθμός των χαρακτήρων που εισήχθησαν. Αν μεταβάλλεται ή όχι θα το αναλυθεί στις επόμενες παραγράφους. **Επίσης θα γίνει προσπάθεια μέσα από τη μελέτη και την εκτέλεση της εφαρμογής να προβλεφθεί, απόλυτα ή κατά πιθανότητα, μέσω του κρυπτογραφήματος η αρχική γλώσσα προέλευσης του.**

3.4.2. Επεξήγηση των κωδικών μέτρησης γραμμάτων κατά την κρυπτογράφηση.

Έπειτα από την εισαγωγή των πακέτων, μέσα στην κλάση JavaApplication1 δηλώνονται οι κρυπταλγόριθμοι κρυπτογράφησης (ecipher) και αποκρυπτογράφησης (dcipher), αρχικοποιείται η έξοδος να είναι "null" και φτιάχνεται ένα αρχείο με όνομα "pososta.txt" στο οποίο θα πηγαίνει η έξοδος του αριθμού εμφάνισης κάθε γράμματος έπειτα από την εκτέλεση της εφαρμογής. Υπάρχει

- ✓ ο δημιουργός της κλάσης `JavaApplication1(SecretKey key)` ο οποίος παίρνει ένα όρισμα,

και επίσης υλοποιούνται:

- ✓ οι μέθοδοι κρυπτογράφησης `public void encrypt(InputStream in, OutputStream out)`,
- ✓ και αποκρυπτογράφησης `public void decrypt(InputStream in, OutputStream out)`.

Ο δημιουργός της κλάσης και οι μέθοδοι που αναφέρθηκαν ήρθαν αυτούσια από τον αλγόριθμο AES και στην ουσία υποδηλώνουν και τον τρόπο λειτουργίας αυτού του αλγορίθμου.

Στην κλάση της main παράγεται ένα κλειδί το οποίο θα χρησιμοποιείται κάθε φορά για την κωδικοποίηση, ενώ επίσης δημιουργούνται τα αρχεία "DESTest.txt" το οποίο λειτουργεί ως είσοδος για την εισαγωγή δεδομένων το οποίο θα είναι και το αρχικό κείμενο προς κρυπτογράφηση, και το "Encrypted.txt", το οποίο αποτελεί αρχείο εξόδου και σε αυτό αποθηκεύεται το κρυπτογραφημένο κείμενο. Υπάρχει επίσης και η αντίστροφη διαδικασία, όπου από το κρυπτογραφημένο κείμενο ανακτάται το αρχικό. Δηλαδή χρησιμοποιείται ως αρχείο δεδομένων το υπάρχον από την προηγούμενη διαδικασία "Encrypted.txt" το οποίο περιέχει το κρυπτογραφημένο κείμενο, και από αυτό ανακτάται το αρχικό το οποίο αποθηκεύεται στο αρχείο εξόδου "Decrypted.txt". Ουσιαστικά, μετά την εκτέλεση της εφαρμογής παρατηρείται ότι όντως το αρχικό κείμενο που υπάρχει στο "DESTest.txt" είναι ίδιο με αυτό από την διαδικασία της αποκρυπτογράφησης, δηλαδή το "Decrypted.txt". Επίσης στην κλάση της main δηλώνονται οι πίνακες γραμμάτων για τα κεφαλαία με όνομα `capital` και για τα πεζά γράμματα με όνομα `small` τόσο στην αγγλική, όσο και στην ελληνική γλώσσα. Δηλώνεται ένας ανιχνευτής με το όνομα "scan", ο οποίος κάνει αναζήτηση στο αρχείο "DESTest.txt".

Πρέπει να σημειωθεί εδώ ότι στον "scan" πρέπει να δοθεί η σωστή διαδρομή όπου βρίσκεται αποθηκευμένο το αρχείο "DESTest.txt". Αν δεν βρεθεί το αρχείο τότε θα εκτυπωθεί το κατάλληλο μήνυμα μη εύρεσης του αρχείου. Στο επόμενο μπλοκ εντολών try, δίνεται η εντολή:

```
output = new BufferedWriter(new FileWriter(file));
```

Δηλώνεται δηλαδή ότι ως έξοδος γράφεται το αρχείο file το οποίο έχει οριστεί παραπάνω ότι είναι το pososta.txt. Φτιάχνεται ένας πίνακας με το όνομα `count` με χωρητικότητας πενήντα (50) θέσεων, 26 θέσεις για τα αγγλικά γράμματα και 24 για τα ελληνικά γράμματα. Όσο ο ανιχνευτής(scan) θα βρίσκει ότι υπάρχει και επόμενος χαρακτήρας, αυτός θα διαβάζεται και θα προστίθεται στο τέλος της γραμμής του αρχείου εξόδου. Δηλώνεται επίσης ο πίνακας χαρακτήρων με όνομα digit ο οποίος μετατρέπει τα γράμματα εισόδου σε χαρακτήρες. Στη συνέχεια γίνεται έλεγχος αν το πλήθος των γραμμάτων του εισερχόμενου κειμένου έχουν εξαντληθεί. Όσο υπάρχουν χαρακτήρες από αυτούς που έχουν δηλωθεί πρωτύτερα στους πίνακες capital και small τότε αυτοί θα προστίθενται αυτοί στο αρχείο εξόδου. Αν δεν βρεθούν άλλοι χαρακτήρες, τότε κλείνει το αρχείο εξόδου.

Διευκρίνηση:

Έχοντας δώσει μία περιγραφή για τον τρόπο λειτουργίας της εφαρμογής, πρέπει επίσης να επεξηγηθεί μία παρατήρηση για το αρχείο εισόδου (DESTest.txt). Σε αυτή την εφαρμογή αναγνώρισης της γλώσσας προέλευσης από κρυπτοκείμενο, σε αντίθεση με την προηγούμενη, δεν επηρεάζει αν στο τέλος των δεδομένων τοποθετηθεί κάποιο σύμβολο, όπως τελεία, θαυμαστικό κτλ. Αυτό που θα γίνει είναι ότι πολύ απλά θα αγνοηθεί ο χαρακτήρας καθώς δεν έχει δηλωθεί στην βάση δεδομένων των χαρακτήρων που έχουν εισαχθεί. Υπενθυμίζεται πως δηλώθηκαν μόνο οι πίνακες κεφαλαίων και πεζών γραμμάτων, αγγλικών και ελληνικών. Για τον ίδιο λόγο, παρατηρείται ότι για την ελληνική γλώσσα δεν μετρώνται τα ελληνικά γράμματα τα οποία τονίζονται. Θεωρούνται διαφορετικοί χαρακτήρες και δεν προστίθενται στο πλήθος κάποιου χαρακτήρα. Αντιθέτως, δεν παίζει ρόλο για ένα γράμμα αν αυτό είναι πρώτο σε μία πρόταση και είναι κεφαλαίο. Τα κεφαλαία και τα πεζά γράμματα θεωρούνται κοινά και αυξάνεται κάθε φορά από τον ανιχνευτή η εύρεσή τους είτε είναι μικρό είτε μεγάλο ως κοινό γράμμα.

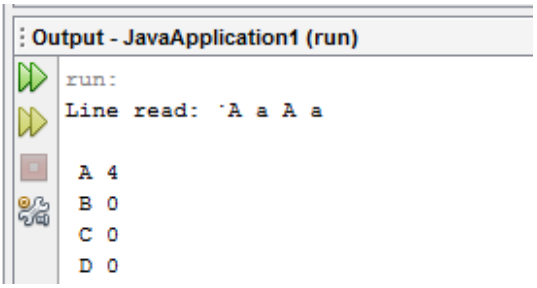
3.5. Πρόβλεψη κειμένου από το ποσοστό γραμμάτων.

Βασιζόμενοι στην εφαρμογή που αναφέρθηκε παραπάνω στην παράγραφο 3.4.2, ***θα παρουσιαστεί μία εκτίμηση για την πρόβλεψη ενός κειμένου, όποιο και αν είναι αυτό, έχοντας όμως ως δεδομένο το κρυπτοκείμενο.*** Δηλαδή γίνεται

προσπάθεια να προβλεφθεί από ένα αρχείο εισόδου, το οποίο είναι ένα κρυπτογραφημένο κείμενο, αν αυτό προέρχεται από την αγγλική ή την ελληνική γλώσσα. Για να γίνει μία τέτοια πρόβλεψη, έχοντας την δυνατότητα να μετρώνται οι χαρακτήρες ενός κρυπτογραφημένου κειμένου θα γίνουν διάφορες μετρήσεις τόσο με αγγλικά όσο και με ελληνικά κρυπτοκείμενα για να εξαχθούν χρήσιμα συμπεράσματα. **Αρχικά θα μελετηθούν τα ποσοστά των γραμμάτων, τα οποία και θα υπολογιστούν στο μήκος του κειμένου έτσι ώστε να τα συγκρίνουμε με επίσημες δημοσιεύσεις, όπως είναι ο ΕΘΕΓ (βλ. κεφάλαιο 1), αν ταυτίζονται ή όχι. Στη συνέχεια θα γίνει η σύγκριση της εμφάνισης ενός γράμματος στο αρχικό κείμενο και στο κρυπτοκείμενο.**

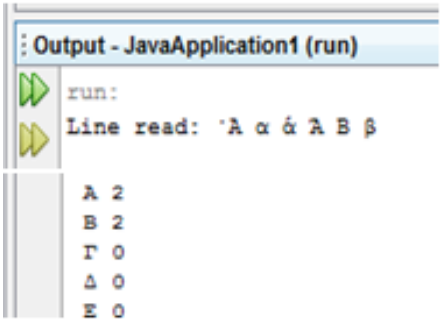
3.5.1. Ανάλυση γραμμάτων.

Έχοντας εκτελέσει την εφαρμογή της αναγνώρισης του κειμένου από το ποσοστό γραμμάτων, δηλαδή από το αρχείο "ptyxiaki", θεωρείται σημαντικό να αναφερθούν κάποια ζητήματα που προέκυψαν κατά την διαδικασία της μελέτης της εφαρμογής. Αρχικά αναφέρεται ότι η εφαρμογή αυτή δίνει την δυνατότητα μέτρησης των γραμμάτων που εισάγονται στο αρχείο εισόδου. Μέσα από την μελέτη για τον τρόπο κρυπτογράφησης των γραμμάτων διαπιστώθηκε πως όλα τα γράμματα προσμετρούνται, είτε είναι πεζά είτε κεφαλαία. Οποιαδήποτε άλλα σύμβολα όπως σημεία στίξης κτλ απλώς παραβλέπονται και συνεχίζεται η ανίχνευση στους επόμενους χαρακτήρες για την ύπαρξη γραμμάτων. Έτσι λοιπόν, στην έξοδο της εφαρμογής υπάρχει συγκεντρωτικά μια λίστα με τα γράμματα, αρχικά τα αγγλικά και έπειτα τα ελληνικά, όπου δίπλα σε κάθε γράμμα υπάρχει ο αριθμός εμφάνισής του στο αρχικό κείμενο. Στον αριθμό αυτό συνυπολογίζονται τα πεζά και τα κεφαλαία δίνοντας τον συνολικό τους αριθμό εμφάνισης. Στην παρακάτω εικόνα απεικονίζεται ενδεικτικά ότι τα πεζά και τα κεφαλαία γράμματα μετρώνται από την εφαρμογή συνολικά.



Εικόνα 3.6: Συνυπολογισμός κεφαλαίων και πεζών γραμμάτων.

Ωστόσο δεν συμβαίνει το ίδιο σε γράμματα τα οποία τονίζονται. Για παράδειγμα, στην εικόνα παρακάτω απεικονίζεται ότι για τα ελληνικά γράμματα, τα φωνήεντα που τονίζονται θεωρούνται διαφορετικοί χαρακτήρες, όπως επίσης και τα κεφαλαία γράμματα. Ενώ δηλαδή υπάρχουν τέσσερα «Α» τα δύο τονίζονται και δεν συνυπολογίζονται στην έξοδο του προγράμματος.



Εικόνα 3.7: Τα τονούμενα δεν συνυπολογίζονται.

Έτσι λοιπόν πρέπει να ληφθεί υπ' όψιν ότι ενδεχομένως ο αριθμός των εμφάνισης των γραμμάτων που θα προκύψει να μην ταυτίζεται απόλυτα με τις μελέτες από πηγές που έχουν δημοσιευτεί (π.χ ΕΘΕΓ).

- ✓ Στην ελληνική γλώσσα, υπάρχουν γράμματα τα οποία τονίζονται. Κάποια από αυτά εμφανίζουν και υψηλό ποσοστό εμφάνισης σε ένα κείμενο. Για παράδειγμα, τα γράμματα "α" και "ε" είναι από τα συνηθέστερα της ελληνικής αλφαβήτου στον γραπτό λόγο. Επειδή όμως από την εφαρμογή αυτή δεν μετρώνται χαρακτήρες οι οποίοι τονίζονται, καθώς θεωρούνται διαφορετικοί χαρακτήρες, ενδεχομένως τα ποσοστά των γραμμάτων αυτών να μην είναι απόλυτα αξιόπιστα. Συνεπώς θα δοθεί ιδιαίτερη προσοχή και βαρύτητα στα σύμφωνα γράμματα της ελληνικής αλφαβήτου.
- ✓ Επίσης από ένα κρυπτοκείμενο παράγονται διάφοροι χαρακτήρες πέραν των γραμμάτων, τα οποία και θα μελετηθούν και θα συγκριθούν μεταξύ των ελληνικών και αγγλικών κρυπτοκειμένων. Ενδεχομένως κάποιος χαρακτήρας να μην εμφανίζεται και σε ελληνικό και σε αγγλικό κρυπτοκείμενο ή αν κάποιος εμφανίζεται πάντοτε σε κάποιο από αυτά τα δύο.
- ✓ Ένα άλλο στοιχείο που θεωρείται σημαντικό προς εξέταση είναι αν μέσα από τα κρυπτοκείμενα υπάρχει κάποιος χαρακτήρας ή γράμμα το οποίο να

επαναλαμβάνεται σε μεγάλο βαθμό, ώστε να φανερώνει έμμεσα και την αρχική γλώσσα προέλευσης του κρυπτοκειμένου.

3.5.2. Συμπεράσματα.

Αρχικά θα παρατεθούν τα αριθμητικά και ποσοστιαία αποτελέσματα που εξήχθησαν με τη βοήθεια του κώδικα της μέτρησης των χαρακτήρων. Η σύγκριση θα γίνει για την ελληνική γλώσσα και θα συγκριθεί με αποτελέσματα του ΕΘΕΓ, αλλά και για την αγγλική και θα εξακριβωθεί αν επιβεβαιώνουν όσα έχουν μελετηθεί από επίσημες βιβλιογραφίες.
Για την αγγλική γλώσσα, χρησιμοποιήθηκε ένα κείμενο 700 λέξεων όπου και έγινε καταμέτρηση των χαρακτήρων από την εφαρμογή. Στον παρακάτω πίνακα παρουσιάζεται ο αριθμός εμφανίσεων και το ποσοστό επί τις εκατό (%) κάθε γράμματος του αγγλικού αλφαβήτου.

Πίνακας 3.2: Μετρήσεις ελληνικών γραμμάτων από την εφαρμογή.

Γράμμα	Εμφανίσεις	Ποσοστό (%)	Γράμμα	Εμφανίσεις	Ποσοστό (%)
A	308	8,57	**N**	251	6,98
B	53	1,47	**O**	258	7,18
C	114	3,17	**P**	66	1,83
D	126	3,50	**Q**	4	0,11
E	470	13,08	**R**	243	6,51
F	80	2,22	**S**	255	7,09
G	69	1,92	**T**	329	9,15
H	174	4,84	**U**	87	2,42
I	267	7,43	**V**	43	1,19
J	7	0,19	**W**	52	1,44
K	20	0,55	**X**	34	0,94
L	155	4,31	**Y**	34	0,94
M	93	2,58	**Z**	1	0,02
Σύνολο				3593	100%

Από τον παραπάνω πίνακα διαπιστώνεται ότι τα γράμματα που παρουσιάζουν υψηλότερο ποσοστό εμφάνισης είναι τα "E", "T", "A", "O" και "I", ενώ αντίστοιχα τα πιο σπάνια είναι τα "Z", "Q", "J", "K" και "X","Y" με ίδιο ποσοστό. Στους παρακάτω πίνακες γίνεται σύγκριση μεταξύ της παρούσας έρευνας και των αποτελεσμάτων από το βιβλίο «*Cipher systems: the protection of communications*» (Beker & Piper, 1982), όπως έχουν παρουσιαστεί στο κεφάλαιο 1.

Πίνακας 3.3: Σύγκριση συχνότερων αγγλικών γραμμάτων.

Γράμμα	Beker & Piper (%)	Ποσοστό στην έρευνά(%)
E	12,70	13,08
T	9,05	9,15
A	8,16	8,57
O	7,50	7,18
I	6,96	7,43

Σύμφωνα με τον παραπάνω πίνακα παρατηρείται ότι τα πρώτα 5 πιο συχνά γράμματα της παρούσας έρευνας είναι αυτά που παρουσιάζουν και οι Beker και Piper με αρκετές προσεγγίσεις και μοναδική διαφορά την σειρά κατάταξης μεταξύ των γραμμάτων "O" και "I". Εν συνεχεία παραθέτονται και τα λιγότερο συχνά γράμματα της αγγλικής γλώσσας.

Πίνακας 3.4: Σύγκριση λιγότερο συχνών αγγλικών γραμμάτων.

Γράμμα	Beker & Piper (%)	Γράμμα	Ποσοστό στην έρευνά μας (%)
Z	0,07	**Z**	0,02
Q	0,09	**Q**	0,11
X	0,15	**J**	0,19
J	0,15	**K**	0,55
K	0,77	**X,Y**	0,94

Σε αυτόν τον πίνακα παρουσιάστηκαν τα λιγότερα χρησιμοποιούμενα αγγλικά γράμματα και παρατηρείται ότι πλησιάζουν αρκετά τα ποσοστά που δημοσίευσαν οι Beker και Piper. Διαπιστώθηκε μια μικρή διαφορά των τιμών που τροποποιούν

την σειρά στην κατάταξή τους, ωστόσο αυτές δεν επηρεάζουν ιδιαίτερα την εξαγωγή συμπερασμάτων.

Για την ελληνική γλώσσα, χρησιμοποιήθηκε ένα κείμενο 500 λέξεων στο οποίο έγινε η καταμέτρηση των χαρακτήρων. Στον παρακάτω πίνακα παρουσιάζονται αναλυτικά ο αριθμός εμφανίσεων και το ποσοστό επί τις εκατό (%) κάθε γράμματος.

Πίνακας 3.5: Μετρήσεις ελληνικών γραμμάτων από την εφαρμογή.

Γράμμα	Εμφανίσεις	Ποσοστό (%)	Γράμμα	Εμφανίσεις	Ποσοστό (%)
Α	307	12,64	**Ν**	191	7,86
Β	14	0,57	**Ξ**	12	0,49
Γ	39	1,64	**Ο**	203	8,36
Δ	49	2,01	**Π**	118	4,85
Ε	205	8,44	**Ρ**	120	4,94
Ζ	4	0,16	**Σ**	129	4,94
Η	87	3,58	**Τ**	207	8,52
Θ	33	1,35	**Υ**	115	4,73
Ι	180	7,41	**Φ**	18	0,74
Κ	104	4,28	**Χ**	42	1,72
Λ	86	3,54	**Ψ**	2	0,08
Μ	111	4,57	**Ω**	54	2,22
Σύνολο				2428	100%

Από τον παραπάνω πίνακα διαπιστώθηκε ότι τα γράμματα που παρουσιάζουν υψηλότερο ποσοστό εμφάνισης είναι τα "Α", "Τ", "Ε", "Ο" και "Ι", ενώ αντίστοιχα τα πιο σπάνια είναι τα "Φ", "Β", "Ξ", "Ζ" και "Ψ". Στους παρακάτω πίνακες παρουσιάζεται η σύγκριση μεταξύ παρούσας έρευνας και των αποτελεσμάτων του ΕΘΕΓ, όπως έχουν παρουσιαστεί στο κεφάλαιο 1.

Πίνακας 3.6: Σύγκριση συχνότερων ελληνικών γραμμάτων.

Γράμμα	Ποσοστό στον ΕΘΕΓ (%)	Ποσοστό στην έρευνά μας (%)
Α	11,49	12,6
Ο	10,14	8,36
Ι	9,32	7,41
Ε	8,65	8,44
Τ	7,98	8,52

Με βάση αυτόν τον πίνακα παρατηρείται ότι τα πρώτα 5 πιο συχνά γράμματα της παρούσας μελέτης είναι αυτά που παρουσιάζει και ο ΕΘΕΓ ως πιο συχνά κατά προσέγγιση. Όπως έχει προαναφερθεί, λόγω του ότι τα περισσότερα συχνά γράμματα είναι φωνήεντα και τονίζονται, ο κώδικάς αυτός τα θεωρεί διαφορετικά σύμβολα και δεν τα συνυπολογίζει. Γι' αυτό το και παρουσιάζονται τα ποσοστά στην παρούσα έρευνα λίγο μικρότερα. Αντίθετα, για το γράμμα "Α" η παρούσα έρευνα το ανέδειξε πιο συχνό από τον ΕΘΕΓ με ποσοστό 12,6 %.

Πίνακας 3.7: Σύγκριση λιγότερο συχνών ελληνικών γραμμάτων.

Γράμμα	Ποσοστό στον ΕΘΕΓ (%)	Ποσοστό στην έρευνά μας (%)
Ψ	0,13	0,08
Ζ	0,35	0,16
Ξ	0,4	0,49
Β	0,68	0,57
Φ	0,74	0,82

Σε αυτόν τον πίνακα παρουσιάστηκαν τα λιγότερα χρησιμοποιούμενα ελληνικά γράμματα και παρατηρείται ότι πλησιάζουν αρκετά τα ποσοστά του ΕΘΕΓ. Αξίζει να αναφερθεί ότι όχι μόνο επαληθεύτηκε η έρευνα του ΕΘΕΓ με μικρές διαφορές των τιμών, αλλά και με την ίδια σειρά προς το λιγότερο συχνό γράμμα. Επίσης πρέπει να αναφερθεί ότι τα πιο σπάνια ελληνικά γράμματα είναι σύμφωνα και δεν τονίζονται και προφανώς για τον συγκεκριμένο λόγο δεν παρουσιάζονται μεγάλες διακυμάνσεις μεταξύ των δύο μετρήσεων.

Συνοψίζοντας, από τις μετρήσεις τις εφαρμογής μας παρουσιάστηκαν οι ποσοστιαίες τιμές των αριθμών εμφανίσεων των γραμμάτων, ελληνικών και αγγλικών. **Η παρούσα έρευνα εξήγαγε αποτελέσματα τα οποία πλησιάζουν κατά πολύ τα αποτελέσματα που έχει παρουσιάσει ο ΕΘΕΓ για την ελληνική γλώσσα, αλλά και οι Beker και Piper για την αγγλική.** Προφανώς οι όποιες διαφορετικές τιμές και μετατοπίσεις των γραμμάτων στις πίνακες που παρουσιάστηκαν, οφείλονται τόσο στο ότι κάποια ελληνικά γράμματα που τονίζονται δεν συνυπολογίζονται από τον κώδικά της παρούσας έρευνας, αλλά και στο ίδιο το περιεχόμενο του κειμένου. Ανάλογα με το ύφος του συγγραφέα ή το είδος ενός κειμένου έχουν παρατηρηθεί και διαφορετικά ποσοστά των γραμμάτων όπως έχει αναφερθεί στο κεφάλαιο 1.

3.5.3. Ανίχνευση κειμένου και σύγκριση της εμφάνισης ενός γράμματος στο αρχικό κείμενο και στο κρυπτοκείμενο.

Όπως έχει διαπιστωθεί και από τα προηγούμενα κεφάλαια, η κρυπτογραφία αποτελεί μία επιστήμη η οποία είναι ιδιαίτερα περίπλοκη στην ανάλυση της. Συνεπώς και η μελέτη των γραμμάτων μέσα σε ένα κρυπτοκείμενο απαιτεί προσεκτική μελέτη.

Στόχος είναι να μελετηθεί ο αριθμός των γραμμάτων σε ένα κρυπτοκείμενο ώστε βασιζόμενοι στην παρούσα έρευνα, να ανιχνευτεί η γλώσσα προέλευσης του κρυπτοκειμένου.

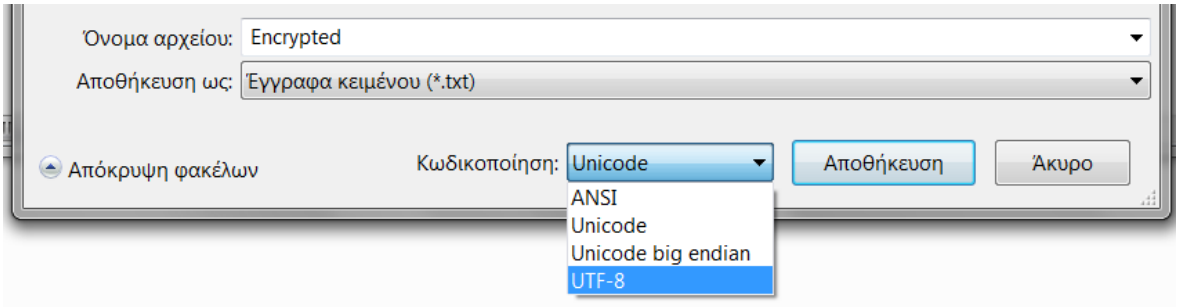
Αρχικά θεωρείται σωστό να επισημανθεί ότι έπειτα από αρκετές εκτελέσεις παρατηρήθηκε πως από την εφαρμογή παράγονται δύο μορφών κρυπτοκείμενα:

1. **₪盂頙錼톈悧賦邦μN摙□좌濧鑃꾉≘부菪潹揲麢겛礽岊垌**
2. **™¬l□%‰TZQ6\-μκτ*l={΄H;vehXnηf%ρ”Nï‡`ƒ{lLn’C**

Στα κρυπτοκείμενα της πρώτης μορφής, όπως παρατηρείται υπάρχουν πολλά σύμβολα τα οποία είναι τύπου Unicode, ιδιαίτερα δυσανάγνωστα και περιέχουν ελάχιστα γράμματα. Τέτοια κρυπτοκείμενα προκύπτουν, όπως διαπιστώθηκε, από αρχεία μεγάλης έκτασης και όχι μιας απλής πρότασης. Επίσης τα αρχεία πολλών χαρακτήρων παράγουν τέτοια κρυπτοκείμενα κυρίως όταν προέρχονται από την αγγλική γλώσσα, σε αντίθεση με την ελληνική όπου τέτοιες εμφανίσεις είναι πιο

σπάνιες. Κατά μεγάλο ποσοστό (90%) ένα κείμενο λίγων χαρακτήρων κρυπτογραφείται με χαρακτήρες της δεύτερης μορφής. Σε αυτά τα κρυπτοκείμενα επίσης γίνεται χρήση των χαρακτήρων Unicode, τα οποία είναι εμφανώς πιο ευανάγνωστα από αυτά της πρώτης μορφής, ωστόσο γίνεται εκτεταμένη χρήση γραμμάτων.

Σημείωση: Οι χαρακτήρες που παρουσιάστηκαν παραπάνω ότι αποτελούν τα κρυπτοκείμενα είναι τύπου Unicode. Αυτά αποθηκεύονται αυτομάτως από την εφαρμογή στο αρχείο "Encrypted.txt". Ο λόγος που η εφαρμογή έχει χωριστεί σε δύο ξεχωριστά προγράμματα οφείλεται στο γεγονός ότι αν ήταν ενιαίο πρόγραμμα δεν θα ήταν δυνατή η ανάγνωση γραμμάτων από ένα κρυπτοκείμενο καθώς θεωρούνται χαρακτήρες Unicode. Εμπειρικά, παρατηρήθηκε ότι οι χαρακτήρες πρέπει να μετατραπούν σε utf-8 ώστε να μπορούν να είναι αναγνώσιμοι. Έτσι λοιπόν, για την διαδικασία της καταμέτρησης των χαρακτήρων από το κρυπτογραφημένο κείμενο, έπειτα από την πρώτη εκτέλεση του προγράμματος όπου μετρώνται απλώς οι χαρακτήρες του κειμένου, το "Encrypted.txt" πρέπει αυτό να μετατραπεί χειρωνακτικά από Unicode σε utf-8 (Εικόνα 3.8).



Εικόνα 3.8:Μετατροπή σε utf-8

Η εκτεταμένη χρήση των γραμμάτων στην δεύτερη μορφή κρυπτοκειμένου, αλλά και το γεγονός ότι τα σύμβολα που χρησιμοποιούνται και στις δύο μορφές είναι ίδια είτε το αρχικό κείμενο είναι αγγλικό είτε ελληνικό, δυσκολεύει κατά πολύ την προσπάθεια να προκύψουν ασφαλή συμπεράσματα.

Αυτό που πρέπει να τονιστεί είναι ότι έπειτα από συγκεκριμένες μετρήσεις με κείμενα ίδιου περίπου αριθμού λέξεων από αγγλικό και ελληνικό κείμενο προέκυψαν κάποια συμπεράσματα τα οποία επιτρέπουν κατά ένα ποσοστό να προβλεφθεί η αρχική προέλευση της γλώσσας ενός κρυπτοκειμένου.

Συγκεκριμένα η ανάλυση έδειξε τα εξής:

Έπειτα από 20 εκτελέσεις του κώδικα ενός αγγλικού κειμένου, όπου και παρήχθη κρυπτοκείμενο της πρώτης μορφής, προέκυψε το συμπέρασμα ότι ένα αγγλικό κρυπτοκείμενο έχει μια τάση στο να εμφανίζει κάποια γράμματα σε ποσοστό μεγαλύτερο από τα υπόλοιπα. ***Αυτά είναι, το "A" αγγλικό και "A" ελληνικό με ποσοστό 25%, το γράμμα "J" και "S" με ποσοστό 20%. τα υπόλοιπα γράμματα της ελληνικής και αγγλικής αλφαβήτου χρησιμοποιούνται με μικρότερα ποσοστά. Μάλιστα κάποια γράμματα εμφανίζουν μηδενικό ποσοστό όπως τα "D", "N", "O","R", "Y" και τα ελληνικά "Γ","Ρ" και "Φ".***

Κατά την διάρκεια των εκτελέσεων παρήχθησαν και κρυπτοκείμενα της δεύτερης μορφής. Σε αυτά υψηλό ποσοστό εμφάνισης είχαν τα αγγλικά γράμματα "J", "L", "P" και "V" και τα ελληνικά "Ε", "Χ" και "Ω".

Στα κρυπτοκείμενα της δεύτερης μορφής, οι χαρακτήρες που αποτελούν το κρυπτοκείμενο είναι είτε σύμβολα σε μορφή Unicode, είτε γράμματα της ελληνικής και αγγλικής αλφαβήτου τα οποία και είναι το ζητούμενο να καταγραφούν. Σε αυτά λόγω και της ύπαρξης μεγάλου αριθμού γραμμάτων προσεγγίστηκαν κάποιες υποτιθέμενες συμπεριφορές του αλγορίθμου. Σε περίπτωση επαλήθευσής τους θα διακρινόταν κάποια τάση να παρουσιάζει ο αλγόριθμος για τα αγγλικά ή τα ελληνικά κρυπτοκείμενα, *ωστόσο αυτές οι προσπάθειες δεν έδειξαν κάποια αξιοπρεπή αποτελέσματα.* Τέτοιες υποθέσεις ήταν:

1. Σύγκριση των συμβόλων Unicode από διαφορετική γλώσσα προέλευσης κρυπτοκειμένου.
2. Σε ένα κρυπτοκείμενο αν υπάρχει μεγαλύτερο πλήθος αγγλικών γραμμάτων ή ελληνικών.
3. Διαφορά της μέγιστης και της ελάχιστης τιμής που μπορεί να παρουσιάζουν τα γράμματα.
4. Πόσα γράμματα έχουν αριθμό εμφάνισης μεγαλύτερο ή μικρότερο από τον μέσο όρο του αριθμού εμφάνισης των γραμμάτων στο σύνολο.

Εν μέρη κάποιες από τις παραπάνω υποθέσεις θα μπορούσαν ίσως να ισχύουν, ωστόσο οι πολλές εκτελέσεις της εφαρμογής απέδειξαν ότι δεν ισχύουν. **Αυτό που μπορεί γενικά να προσεγγιστεί είναι το γράμμα που έχει τον μεγαλύτερο αριθμό εμφανίσεων. Κάτι τέτοιο όμως μπορεί να αποδειχθεί με μικρό ποσοστό βεβαιότητας.** Συγκεκριμένα για κρυπτοκείμενο της δεύτερης μορφής **αν μεγαλύτερο αριθμό εμφανίσεων έχει το γράμμα "G", "I","O" ή**

κάποιο από τα ελληνικά "Π", "Σ", "Τ" τότε αυτό προέρχεται από την ελληνική γλώσσα. Αν το μεγαλύτερο αριθμό εμφανίσεων παρουσιάζουν τα γράμματα "J", "L", "P", "Q", "V", και τα ελληνικά "Ε", "Ν", "Χ" και "Ω" τότε αυτό πιθανόν να προέρχεται από την αγγλική γλώσσα. Σε περίπτωση μη εμφάνισης κανενός γράμματος, τότε δεν μπορεί να τεκμηριωθεί κάποιο συμπέρασμα.

Συνοπτικά: Ανάλογα από τον τύπο του κρυπτοκειμένου που συναντάται, χρησιμοποιούνται διαφορετικά κριτήρια για να ανιχνευτεί κατά πιθανότητα η αρχική γλώσσα προέλευσης του κρυπτοκειμένου. Αν δηλαδή σε ένα κρυπτοκείμενο είναι εύκολα αντιληπτό πως υπάρχουν πολλά γράμματα της ελληνικής και της αγγλικής αλφαβήτου, τότε θα προσεχθεί το γράμμα εκείνο που παρουσιάζει τον μεγαλύτερο αριθμό εμφάνισης στο κρυπτοκείμενο. Αν όμως στο κρυπτοκείμενο τα γράμματα είναι ελάχιστα και αντικρίζεται σε μεγάλο βαθμό σύμβολα, τότε θα γίνουν υποθέσεις με βάση την εμφάνιση συγκεκριμένων γραμμάτων.

Συνεπώς με απόλυτη βεβαιότητα αναφέρεται ότι η κρυπτογραφία μεταβάλλει σε μεγάλο βαθμό τον αριθμό εμφάνισης των γραμμάτων μεταξύ του αρχικού κειμένου και ενός κρυπτοκειμένου.

3.6. ΕΠΙΛΟΓΟΣ

Στο κεφάλαιο αυτό έγινε αναφορά στις εφαρμογές λογισμικού που αναπτύχθηκαν από τους φοιτητές της πτυχιακής εργασίας και παρουσιάστηκαν τα συμπεράσματα που εξήχθησαν από αυτές. Παρουσιάστηκε μια εφαρμογή η οποία έχει την δυνατότητα να κρυπτογραφεί ένα κείμενο και βασιζόμενη σε αυτό να μπορεί να προβλέπει την αρχική γλώσσα του κειμένου.

Επίσης αναπτύχθηκε μία εφαρμογή η οποία περιλαμβάνει όλη την διαδικασία της κρυπτογράφησης ενός κειμένου με την χρήση του αλγορίθμου AES. Η δεύτερη αυτή εφαρμογή έχει την δυνατότητα καταμέτρησης των γραμμάτων τόσο στο αρχικό κείμενο όσο και στο κρυπτοκείμενο.

Αποδείχθηκε ότι η κρυπτογραφία επηρεάζει, και συγκεκριμένα μεταβάλλει σε μεγάλο βαθμό τον αριθμό εμφάνισης των γραμμάτων.

Τέλος, βασιζόμενοι σε αυτό το λογισμικό, μπορεί κατά πιθανότητα να προβλεφθεί η αρχική γλώσσα προέλευσης ενός κρυπτοκειμένου.

Στο επόμενο κεφάλαιο θα γίνει αναφορά στους κωδικούς πρόσβασης ώστε να εξεταστεί μετέπειτα η κατανομή των γραμμάτων σε κωδικούς πρόσβασης από Έλληνες χρήστες στα ηλεκτρονικά μέσα.

ΚΕΦΑΛΑΙΟ 4

4. ΚΑΤΑΝΟΜΗ ΓΡΑΜΜΑΤΩΝ ΚΑΙ ΚΩΔΙΚΟΙ ΠΡΟΣΒΑΣΗΣ

4.1. ΠΡΟΛΟΓΟΣ

Σε αυτό το κεφάλαιο γίνεται μία περιγραφή για τους κωδικούς πρόσβασης (passwords), αναφέροντας την αναγκαιότητά τους στη σύγχρονη εποχή, κυρίως αυτών του διαδικτύου. Οι κωδικοί αποτελούν ένα εύκολο μέσο ασφαλείας των πληροφοριών και άλλων προσωπικών δεδομένων με αποτέλεσμα να έχουν αυξηθεί οι απόπειρες παραβίασης τους. Παραθέτοντας κάποιες έρευνες θα αναφερθεί η βαρύτητα που δίνουν οι απλοί χρήστες κωδικών, στους τρόπους δημιουργίας τους, την ασφάλειά τους και τις ενδεχόμενες επιπτώσεις από κάποια παραβίασή τους. Επιπλέον, στα πλαίσια της πτυχιακής εργασίας διανεμήθηκε σχετικό ερωτηματολόγιο για την εξαγωγή συμπερασμάτων, σε διάφορους χρήστες, για την πολυπλοκότητα των κωδικών πρόσβασης, το μέγεθός τους και την κατανομή των γραμμάτων σε αυτούς. Τέλος, προβάλλονται κωδικοί πρόσβασης οι οποίοι θεωρούνται αδύναμοι, αλλά παρέχονται και κωδικοί πρόσβασης (βασισμένοι στην έρευνα) τα οποία θα πρέπει να αποφεύγουν οι Έλληνες χρήστες.

4.2. ΕΙΣΑΓΩΓΙΚΑ ΓΙΑ ΤΟΥΣ ΚΩΔΙΚΟΥΣ ΠΡΟΣΒΑΣΗΣ

Ο **κωδικός πρόσβασης (password)** είναι μία μυστική λέξη ή συμβολοσειρά χαρακτήρων που χρησιμοποιείται για έλεγχο ταυτότητας, για να αποδείξει την ταυτότητα μας ή για να προσδώσει πρόσβαση σε μία εφαρμογή. Ένας κωδικός πρόσβασης θα πρέπει να τηρείται μυστικός και να μην αποκαλύπτεται σε άλλα άτομα στα οποία δεν επιτρέπεται η πρόσβαση (Summers and Bosworth, 2004).
Ένας άλλος ορισμός για τον **κωδικό πρόσβασης** είναι ότι αποτελεί μία από τις περισσότερο χρησιμοποιούμενες μεθόδους από τα συστήματα αυθεντικοποίησης. Είναι ακολουθίες χαρακτήρων που πρέπει να θυμούνται οι χρήστες προκείμενου να προστατεύσουν τον υπολογιστή τους ή άλλες εφαρμογές από παραβιάσεις (Ives, et. al, 2004).

Επίσης η έννοια του **κωδικού πρόσβασης** ορίζεται από την microsoft (‘microsoft-what is a password’), ως «μία συμβολοσειρά χαρακτήρων που μπορούν να χρησιμοποιούν οι χρήστες για να συνδεθούν σε έναν υπολογιστή και να έχουν πρόσβαση σε αρχεία, προγράμματα και άλλους πόρους. Οι κωδικοί πρόσβασης μπορούν να διασφαλίσουν ότι η προσπέλαση στον υπολογιστή επιτρέπεται μόνο σε άτομα που έχουν εξουσιοδοτηθεί για αυτό το σκοπό».

Η χρήση του κωδικού πρόσβασης δεν είναι κάτι καινούριο στην κοινωνία μας. Σε στρατιωτικές μονάδες ήταν απαραίτητη η χρήση τους για την είσοδο σε κάποια περιοχή. Αυτό γινόταν με την ανταλλαγή και επιβεβαίωση ενός *συνθηματικού*, μεταξύ φρουρών και ατόμων που επιθυμούν να εισέλθουν σε μία συγκεκριμένη περιοχή. Στην σύγχρονη εποχή, το όνομα ενός χρήστη και ο κωδικός πρόσβασης χρησιμοποιούνται συχνά από τα ανθρώπους κατά την διάρκεια της καταγραφής τους σε μία διαδικασία η οποία ελέγχει την πρόσβαση σε προστατευόμενα υπολογιστικά λειτουργικά συστήματα, κινητά τηλέφωνα, αυτόματες ταμειακές μηχανές (ATM) κτλ. Ένας τυπικός χρήστης ενός υπολογιστή μπορεί να απαιτεί κωδικούς πρόσβασης για διάφορους σκοπούς: σύνδεση των λογαριασμών του υπολογιστή, την ανάκτηση email από διακομιστές, πρόσβαση σε προγράμματα βάσεις δεδομένων, ιστοσελίδες κ.α.

Παρά το όνομά του, δεν είναι υποχρεωτικό ο κωδικός πρόσβασης να είναι μία πραγματική λέξη. Αντιθέτως, μια λέξη password η οποία δεν αποτελεί πραγματική λέξη καθιστά ακόμη πιο δύσκολο να προβλεφθεί ο κωδικός αυτός. Κάποιοι κωδικοί πρόσβασης αποτελούνται από πολλαπλές λέξεις και για αυτόν τον λόγο καλούνται και ως "συνθηματική φράση" (pass phrase). Ο όρος **κωδικός** χρησιμοποιείται μερικές φορές όταν η μυστική αυτή πληροφορία του κωδικού αποτελείται μόνο από αριθμούς, όπως για παράδειγμα το PIN που συχνά χρησιμοποιείται στα μηχανήματα αυτόματης ανάληψης.

Για λόγους πληρέστερης επιβεβαίωσης της ταυτότητας μας από μία υπολογιστική συσκευή σε μία άλλη, οι κωδικοί πρόσβασης εμφανίζουν κάποια σημαντικά μειονεκτήματα (όπως το να κλαπούν, να ξεχαστούν κ.α.). Υπάρχουν όμως συστήματα ασφαλείας, που βασίζονται σε συστήματα κρυπτογραφικών πρωτοκόλλων και τα οποία είναι αρκετά δύσκολο να παρακαμφτούν.

4.3. ΔΗΜΙΟΥΡΓΙΑ ΚΑΙ ΑΣΦΛΑΕΙΑ ΚΩΔΙΚΩΝ ΠΡΟΣΒΑΣΗΣ

4.3.1. Αναγκαιότητα κωδικών πρόσβασης.

Οι περισσότεροι οργανισμοί προσπαθούν να προστατέψουν τα συστήματά τους από μη εξουσιοδοτημένη πρόσβαση, συνήθως μέσω της χρήσης των κωδικών πρόσβασης. Παρά την χρήση όλο και πιο εξελιγμένων μηχανισμών αυθεντικοποίησης, ο αριθμός των προβλημάτων ασφαλείας εξακολουθεί να αυξάνεται (Gordon, 1995; Hitchings, 1995). Μία μη εξουσιοδοτημένη πρόσβαση σε ένα σύστημα, η οποία θα συνεπάγεται την απόκτηση πληροφοριών ή κακή χρήση του συστήματος, συνήθως οφείλεται στο «σπάσιμο» (cracking) των κωδικών ασφαλείας του χρήστη από τους «hackers». Έρευνες έχουν αποδείξει ότι οι μηχανισμοί ασφαλείας είναι λιγότερο αποδοτικοί από ότι γενικά θεωρούνται (Hitchings, 1995).

Η **εμπιστευτικότητα** θεωρείται ως απαραίτητο στοιχείο στην ασφάλεια της πληροφορίας και μαζί με την αυθεντικοποίηση του χρήστη, ο κύριος μηχανισμός απόκτησής της (Parker, 1992). Η διαδικασία της αυθεντικοποίησης μπορεί να διακριθεί σε δύο διαφορετικά στάδια: **την ταυτότητα χρήστη** (user identification, User ID) και τον **κωδικό πρόσβασης** (password). Η ταυτότητα του χρήστη αρχικά αλληλεπιδρά με το σύστημα έτσι ώστε να προσδιορίσει ποιος είναι ο χρήστης, ωστόσο αυτό μόνο δεν αρκεί για να προσδώσει ασφάλεια. Στο δεύτερο στάδιο της αυθεντικοποίησης, απαιτείται η επιβεβαίωση του χρήστη δηλαδή ότι είναι ο νόμιμος χρήστης που αναφέρθηκε στο πρώτο στάδιο. Σε αυτό το σημείο είναι καθοριστική η χρήση του κωδικού πρόσβασης γιατί αυτός μπορεί να προσδιορίσει τον νόμιμο χρήστη και για αυτό τον λόγο ο κωδικός πρέπει να τηρείται μυστικός (Groza and Petrica, 2005).

Γενικά, οι κωδικοί πρόσβασης δημιουργούνται από ένα σύστημα με τις λεγόμενες γεννήτριες κωδικού πρόσβασης (system-generated passwords) που επιβεβαιώνει ότι ο χρήστης χρησιμοποιεί έναν σχετικά ασφαλή συνδυασμό χαρακτήρων. Ωστόσο, αρκετοί χρήστες θεωρούν δύσκολο το να θυμούνται τους κωδικούς πρόσβασης που δημιουργούνται και γι' αυτό συνηθίζουν να τους διατηρούν κάπου χειρόγραφα. Επιπλέον, έχουν εντοπιστεί κίνδυνοι ασφαλείας από αυτά τα συστήματα δημιουργίας των κωδικών πρόσβασης (Kapa, et. al., 2007; Wiedenbeck, et. al, 2005). Οι δύο παραπάνω λόγοι έχουν οδηγήσει τους χρήστες

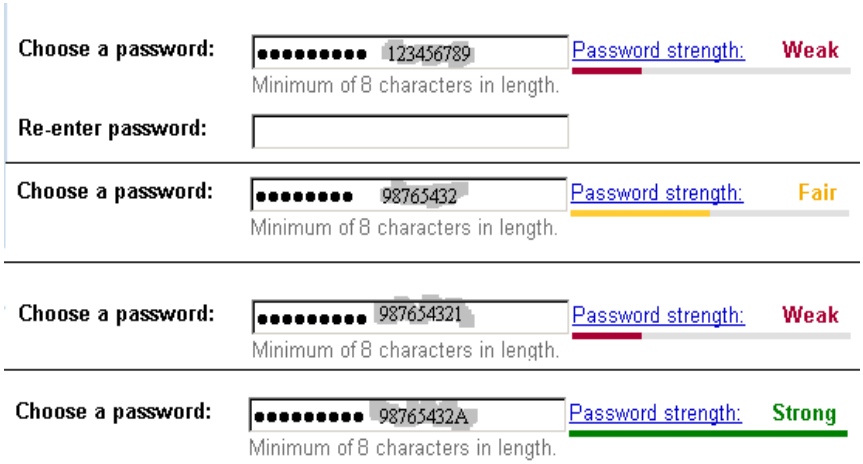
να αποφεύγουν την χρήση αυτών των συστημάτων και να προσπαθούν οι ίδιοι να επινοήσουν password (user-generated passwords). Πέρα από τους κωδικούς μιας λέξης (one-word password), υπάρχουν και άλλοι μηχανισμοί αυθεντικοποίησης που χρησιμοποιούνται (Adams, et. al., 1997), όπως:

- Συνθηματική φράση (passphrase): Απαιτούμενη φράση αντί μιας λέξης.
- Συγγενικοί κωδικοί πρόσβασης (cognitive passwords): Προσωπικές πληροφορίες με τη μορφή ερώτησης-απάντησης.
- Συνδεόμενοι κωδικοί πρόσβασης (associated passwords): Μία ακολουθία λέξεων και συσχετίσεων.
- Προσωπικοί αριθμοί ταυτότητας (PINs).

4.3.2. Ανθεκτικότητα κωδικού πρόσβασης.

Η ανθεκτικότητα του κωδικού πρόσβασης είναι ένα μέτρο που προσδιορίζει την αποτελεσματικότητα του κωδικού πρόσβασης στο να αντιστέκεται και να καθίσταται δύσκολος και ασφαλής σε παραβιάσεις του από κακόβουλους χρήστες. Είναι επίσης μία γενική δομή, η οποία εκτιμά πόσες πιθανές εισόδους θα έπρεπε να δοκιμάσει κάποιος ώστε να καταφέρει να εισέλθει στον λογαριασμό, κατά μέσο όρο, ή να το μαντέψει σωστά. Η ανθεκτικότητα του κωδικού προκύπτει ως μία συνάρτηση κάποιων παραγόντων, όπως το μέγεθος, η πολυπλοκότητά του και η τυχαία διάταξή του (Gehringer, 2002).

Συχνά σε διάφορες εφαρμογές όπου ζητείται η δημιουργία ενός κωδικού, υπάρχει είτε κάποιος περιορισμός όσον αφορά το πλήθος των χαρακτήρων που το αποτελούν είτε εμφανίζεται αυτόματα κατά την δημιουργία του το αντίστοιχο μήνυμα σχετικά με την ανθεκτικότητά του. Στον παρακάτω εικόνα απεικονίζεται η μορφή αυτής της μεθοδολογίας.



Εικόνα 4.1: Απεικόνιση της ανθεκτικότητας ενός κωδικού πρόσβασης.

Σύμφωνα με το Ομοσπονδιακό Πρότυπο Επεξεργασίας Πληροφοριών των Ηνωμένων Πολιτειών (US Federal Information Processing Standards-FIPS) (‘Federal Information Processing Standards’, 2001) υπάρχουν αρκετά κριτήρια που μπορούν να χρησιμοποιηθούν για να βεβαιώσουν τα διαφορετικά επίπεδα ασφάλειας των password. Συγκεκριμένα, η *σύνθεση ενός κωδικού,* σχετίζει άμεσα το επίπεδο ασφάλειας του κωδικού με το μήκος των χαρακτήρων που το απαρτίζουν (Adams et. al., 1997). Ένας αλφαριθμητικός κωδικός πρόσβασης είναι συνεπώς πιο ασφαλές από έναν άλλο που αποτελείται μόνο από γράμματα. Επίσης, οι κωδικοί πρόσβασης *μικρής διάρκειας* (short password lifetime), αυτά δηλαδή που αλλάζονται έπειτα από σύντομο χρονικό διάστημα, θεωρούνται ότι μειώνουν τον κίνδυνο που σχετίζεται με την ανυποψίαστη αθέμιτη χρήση ενός κωδικού πρόσβασης μιας λέξης. Επίσης, σύμφωνα με το FIPS, η μυστικότητα του κωδικού πρόσβασης θεωρείται απαραίτητη για την ασφάλειά του.

Υπάρχουν έρευνες που αποδεικνύουν ότι πολλοί χρήστες δεν ακολουθούν μεθόδους μιας ασφαλούς δημιουργίας ενός κωδικού. Συγκεκριμένα, σε μια έρευνα που έκανε ο DeAlvare (De Alvaré, 1988) ανακάλυψε ότι καθώς ένας κωδικός πρόσβασης επιλεχθεί από ένα χρήστη, τότε είναι απίθανο αυτό να αλλαχθεί όσο ο χρήστης είναι ικανοποιημένος από αυτό (Spafford, 1992). Η έρευνα του DeAlvare συνεχίστηκε έως το 1990 για να αποδείξει ότι αν είχαν την δυνατότητα οι χρήστες, από τις εφαρμογές που χρησιμοποιούν θα προτιμούσαν να φτιάξουν ένα κωδικό με όσο το δυνατόν λιγότερους χαρακτήρες. Σύμφωνα με την έρευνα δεν υπάρχει περίπτωση αμφισβήτησης αυτής της άποψης, απλώς αυτή η συμπεριφορά οφείλεται στο ότι οι χρήστες είναι απρόσεκτοι. Τα τμήματα ασφαλείας των

διαφόρων οργανισμών προσπαθούν να αλληλεπιδράσουν με τους χρήστες με σκοπό να τους αλλάξουν αυτή την συμπεριφορά μέσω διάφορων συστημάτων, όπως μηχανισμούς λήξης ενός κωδικού πρόσβασης, περιορισμούς κατά την δημιουργία του password. Μέσω αυτών των μηχανισμών θεωρείται πως οι χρήστες θα συμμορφωθούν με τους νέους κανόνες ασφαλείας και έτσι θα μειωθεί η εμφάνιση τέτοιας συμπεριφοράς. Σκοπός είναι να απαντήσουν στα προβλήματα ασφαλείας προωθώντας περισσότερο περιοριστικές πολιτικές αυθεντικότητας όπως:

- Συχνότερη αλλαγή κωδικών (π.χ. κάθε μήνα).
- Μεγαλύτεροι και περισσότερο πολύπλοκοι κωδικοί.
- Μείωση του ποσοστού των επιτρεπόμενων λαθών κατά την εισαγωγή.

Παρά την προσπάθεια που έγινε από τα τμήματα ασφαλείας, αυτοί οι μηχανισμοί δεν είχαν το επιθυμητό αποτέλεσμα. Ενώ αναμενόταν ότι οι χρήστες θα συμμορφώνονταν με τους νέους κανόνες ασφαλείας κάτι τέτοιο όχι μόνο δεν συνέβη αλλά είχε και το αντίθετο αποτέλεσμα. Αποδείχθηκε ότι **όσο περισσότερο περιοριστικοί και πολύπλοκοι μηχανισμοί υπάρχουν, τόσο περισσότερο οι χρήστες τους παρακάμπτουν**. Αν και αυτό είναι παράδοξο, ο λόγος είναι ότι οι περιοριστικοί αυτοί μηχανισμοί δημιουργούν περισσότερα προβλήματα κατά την χρήση τους. Έχει διατυπωθεί ότι τα χαρακτηριστικά τα οποία καθιστούν ένα κωδικό περισσότερο ασφαλή, ταυτόχρονα τον κάνουν και πιο δύσκολο στην απομνημόνευσή του (Carroll, 1996). Σχετικά με αυτό θέμα έχουν γίνει προσπάθειες να αναπτυχθούν μηχανισμοί οι οποίοι θα παράγουν κωδικούς πιο εύκολους στη απομνημόνευσή τους (Barton Marthalee and Ben, 1984). Ωστόσο, όλες αυτές οι προτάσεις φαίνεται πως είναι περιορισμένες καθώς δεν υπάρχει το ανάλογο ενδιαφέρον από τους χρήστες.

4.3.3. Επιλογή ασφαλούς κωδικού.

Σχετικά με την επιλογή ενός ασφαλούς κωδικού έχουν αναπτυχθεί διάφορες μελέτες και έρευνες. Είναι πλέον γνωστό πως **υπάρχουν στρατηγικές και μέθοδοι για να κατασκευάσουμε έναν ανθεκτικό κωδικό.** Μπορεί να θεωρείται από κάποιους ότι ίσως είναι μία δύσκολη διαδικασία όσον αφορά την πολυπλοκότητα ενός κωδικού που πρέπει να δημιουργήσουμε, ειδικότερα όταν

σχετίζεται με εφαρμογές οι οποίες έχουν ιδιαίτερη σημασία στο να προστατεύσουν προσωπικά δεδομένα. Ωστόσο η επιλογή ενός ισχυρού κωδικού στην ουσία αποτελεί μία πάρα πολύ απλή διαδικασία, αρκεί πάντοτε να γνωρίζουν οι χρήστες τι πρέπει να αποφεύγουν και πως μπορούν να γίνουν περισσότερο δημιουργικοί.
Έτσι λοιπόν παραθέτεται η διαδικασία δημιουργίας ενός ασφαλούς και ανθεκτικού κωδικού λαμβάνοντας υπ' όψιν τα παρακάτω (Kuo et. al., 2006):

1. **Φτιάξτε τον κωδικό σας όσο το δυνατόν πιο μεγάλο σε χαρακτήρες.** Όσο μεγαλύτερος είναι τόσο δυσκολότερο είναι να παραβιαστεί σε κάποια επίθεση. Πάντοτε να χρησιμοποιούνται τουλάχιστον έξι χαρακτήρες, εκ των οι δύο να είναι αριθμοί.
2. **Πρέπει να χρησιμοποιείτε όσο το δυνατόν διαφορετικούς χαρακτήρες όταν διαμορφώνετε τον κωδικό σας.** Χρησιμοποιήστε αριθμούς, σημεία στίξεως και μίξη μικρών και κεφαλαίων γραμμάτων έτσι ώστε να τον κάνετε πιο ανθεκτικό.
3. **Αποφύγετε να χρησιμοποιήσετε προσωπικές σας πληροφορίες** στον κωδικό σας ώστε κάποιος να μην μπορεί να τον μαντέψει για εσάς.
4. **Αποφύγετε τις λέξεις του λεξικού, σε οποιαδήποτε γλώσσα.** Οι εγκληματίες χρησιμοποιούν εξελιγμένα εργαλεία που μαντεύουν γρήγορα τους κωδικούς πρόσβασης που βασίζονται σε λέξεις πολλών λεξικών, συμπεριλαμβανομένων λέξεων γραμμένων ανάποδα, συνηθισμένων ανορθογραφιών και αντικαταστάσεων. Σε αυτές περιλαμβάνονται όλων των ειδών οι βρισιές και οι βλασφημίες.
5. **Ποτέ μην χρησιμοποιείται ως password σε κάποια εφαρμογή έναν κωδικό ο οποίος ταυτόχρονα αποτελεί και κωδικό μιας άλλης εφαρμογής**, όπως ενός τραπεζικού λογαριασμού (PIN).
6. **Μην χρησιμοποιείται μόνο αριθμούς ή μόνο γράμματα.**
7. **Κάντε συνδυασμούς γραμμάτων και αριθμών** παρεμβάλλοντας ανάμεσά τους σύμβολα ή σημεία στίξης (π.χ κάποιο από τα: ` ! " ? $? % ^ & * ( ) _ - + = { [}] : ; @ ' ~ # | \ < , > . ? /).
8. **Μην χρησιμοποιείτε passwords τα οποία είναι εύκολο να διαμορφωθούν κατά την πληκτρολόγηση**. **Αποφύγετε τις ακολουθίες ή τους επαναλαμβανόμενους χαρακτήρες.** Τα "12345678", "222222", "abcdefg" ή τα γράμματα που γειτονεύουν στο πληκτρολόγιο δεν χρησιμεύουν για τη δημιουργία ισχυρών κωδικών. Δηλαδή passwords

123456, qwerty και γενικότερα γράμματα τα οποία βρίσκονται σε διπλανή θέση στο πληκτρολόγιο θα πρέπει να αποφεύγονται.

9. **Προσθέστε πολυπλοκότητα** συνδυάζοντας κεφαλαία και πεζά γράμματα και αριθμούς. Είναι χρήσιμο να χρησιμοποιήσετε εναλλαγές γραμμάτων ή ανορθογραφίες. Για παράδειγμα, στην παρακάτω κωδική φράση, "My son Aiden is three years old", γράψτε ανορθόγραφα το όνομα του Aiden ή αντικαταστήστε τη λέξη "τριών/three" με τον αριθμό 3. Υπάρχουν πολλές διαφορετικές αντικαταστάσεις και, όσο μεγαλύτερη είναι η φράση, τόσο πιο περίπλοκος μπορεί να γίνει ο κωδικός. Μπορεί δηλαδή η κωδική φράση να γίνει "My SoN Ayd3N is 3 yeeRs old". Αν ο υπολογιστής ή το διαδικτυακό σύστημα δεν υποστηρίζει κωδική φράση, χρησιμοποιήστε την ίδια τεχνική στον μικρότερο κωδικό πρόσβασης. Έτσι θα προκύψει ο κωδικός πρόσβασης "MsAy3yo".
10. **Αντικαταστήστε ορισμένους ειδικούς χαρακτήρες.** Μπορείτε να χρησιμοποιήσετε σύμβολα που μοιάζουν με γράμματα, να συνδυάσετε λέξεις (αφαιρέστε τα διαστήματα) και άλλους τρόπους να κάνετε τον κωδικό πρόσβασης πιο περίπλοκο. Με αυτές τις τεχνικές δημιουργείται η κωδική φράση: "MySoN 8N i$ 3 yeeR$ old" ή η κωδική φράση (με το πρώτο γράμμα της κάθε λέξης) "M$8ni3y0".
11. **Ελέγξτε τον νέο σας κωδικό πρόσβασης με το Password Checker.** Το Password Checker είναι μια δυνατότητα αυτής της τοποθεσίας Web που δεν κάνει καταγραφές και σας βοηθά να προσδιορίσετε πόσο ισχυρός είναι ο κωδικός πρόσβασης που δημιουργείται, καθώς αυτός πληκτρολογείται.
12. **Αποφύγετε τη χρήση διαδικτυακών εργαλείων αποθήκευσης.** Αν οι κακόβουλοι χρήστες βρουν αυτούς τους κωδικούς πρόσβασης αποθηκευμένους διαδικτυακά ή σε δικτυωμένο υπολογιστή, έχουν πρόσβαση σε όλες τις πληροφορίες.

Σε γενικές γραμμές, έχει παρατηρηθεί ότι οι χρήστες δεν δίνουν την ανάλογη προσοχή σε αυτές τις απλές οδηγίες ή τις αγνοούν πλήρως. Είναι γενικώς κατανοητό ότι το να δημιουργηθεί ένας κωδικός από έναν χρήστη, μπορεί να έχει μεγάλη ασφάλεια, καθώς θα γίνεται πολύπλοκος, αλλά αυτό δεν συνεπάγεται απόλυτα και την ασφάλειά του. Κι αυτό οφείλεται στο γεγονός ότι ένας πολύπλοκος κωδικός ή γενικότερα οι συχνές αλλαγές τέτοιων κωδικών δεν συντελούν στην εύκολη απομνημόνευσή τους. Ως αποτέλεσμα, οι χρήστες

συνηθίζουν, ενώ διαθέτουν ένα πολύ ισχυρό κωδικό πρόσβασης να τον σημειώνουν χειρόγραφα και τον αποθηκεύουν κάπου κοντά στον υπολογιστή τους. Είναι προφανές ότι αυτό δεν εγγυάται την ασφάλειά του κωδικού από άλλους χρήστες που θα πέσει εύκολα στα χέρια τους. Συνέπειες των αδύναμων κωδικών ή περιπτώσεις κλοπής του κωδικού πρόσβασης θα αναλυθούν στις επόμενες παραγράφους.

Συνοψίζοντας, μια καλή συμβουλή για τον κωδικό πρόσβασης είναι να μην χρησιμοποιείται ποτέ μια πραγματική λέξη ή μέρος κάποιας λέξης. Με την χρήση απλών εργαλείων μπορεί κανείς να "σπάσει" τον κωδικό πολύ εύκολα και σε σύντομο χρονικό διάστημα, αν για παράδειγμα υπάρχει ως κωδικός πρόσβασης η λέξη "kalhmera", σε αντίθεση με ένα κωδικό της μορφής "f1djt&9B", αν και όπως παρατηρείται και οι δύο κωδικοί αποτελούνται από οχτώ χαρακτήρες. Αυτό που έχει απλώς να κάνει ο χρήστης ώστε να φτιάξει ισχυρό τον κωδικό του πρόσβασης είναι απλώς να το επεκτείνει σε μέγεθος χαρακτήρων. Όσο μεγαλύτερος είναι ένας κωδικός πρόσβασης, τόσο δυσκολότερο είναι αυτός να σπάσει και φυσικά απαιτεί πιο εξειδικευμένους μηχανισμούς και εργαλεία γενικότερα. Δηλαδή, αναφέροντας ένα απλό παράδειγμα, είναι προτιμότερο να χρησιμοποιήσει κάποιος το password 'HlioloustiMeraSimera' από το 'S3cur!ty'. Ακόμη και αν το δεύτερο φαίνεται δύσκολο για απομνημόνευση και χρησιμοποιεί μία ποικιλία χαρακτήρων από αριθμούς, σύμβολα, κεφαλαία και μικρά γράμματα είναι λιγότερο ασφαλές από το πρώτο. Και αυτό συμβαίνει διότι είναι μικρότερο σε μέγεθος και μπορεί να σπάσει με πιο απλά εργαλεία. Ενώ ο πρώτος κωδικός αποτελεί μια εύκολη έκφραση, έχει περισσότερους χαρακτήρες και αυτό το κάνει ακόμη πιο ασφαλές και πιο ανθεκτικό (Yan, et. al., 2000).

Γενικώς, οτιδήποτε χρησιμοποιείται με περισσότερους από εφτά χαρακτήρες είναι ένα καλό ξεκίνημα. Για να επιβεβαιώσει κανείς ότι διαθέτει ένα υψηλό και αξιόπιστο επίπεδο ασφάλειας του κωδικού του πρόσβασης μπορεί να επιλέξει έναν κωδικό με δεκαπέντε χαρακτήρες.

4.3.4. Εγκυρότητα ενός κωδικού.

Ο πιο προφανής τρόπος για να ελέγξει ο χρήστης την εγκυρότητα ενός κωδικού είναι να προσπαθήσει να τον χρησιμοποιήσει στην εφαρμογή που υποτίθεται ότι

χρειάζεται ο κωδικός που προσπαθεί να προβλέψει. Ωστόσο, κάτι τέτοιο θα είναι αργό και προφανώς κάποια συστήματα να καθυστερούν ή ακόμη και να «μπλοκάρουν» την πρόσβαση έπειτα από ένα συγκεκριμένο αριθμό αποτυχημένων προσπαθειών. Επίσης, συστήματα τα οποία χρησιμοποιούν κωδικούς πρόσβασης για την επιβεβαίωση εισόδου σε αυτό πρέπει να τα αποθηκεύουν σε μία βάση δεδομένων για να ελέγξουν την διαθεσιμότητα των τιμών. Μία συνηθισμένη πρακτική είναι να αποθηκεύεται μόνο η συνάρτηση κατακερματισμού ενός κωδικού αντί του κωδικού. Αν αυτή η συνάρτηση είναι αρκετά ισχυρή, τότε είναι πολύ δύσκολο να αποκτηθεί από έναν ανεπιθύμητο χρήστη. Αν όμως έχουν χαθεί αρχεία τα οποία περιέχουν αυτές τις συναρτήσεις, τότε αυτές πλέον θεωρούνται γνωστές από τους άλλους χρήστες και είναι εύκολο να παραβιαστεί ο κωδικός μας.

4.3.5. Αποφυγή στοιχείων στους κωδικούς πρόσβασης.

Για να γίνει κατανοητό και πιο σαφές τι θα πρέπει να αποφεύγουν οι καθημερινοί χρήστες που χρησιμοποιούν σε διάφορες εφαρμογές, όπως αναφέρουν οι Garfinkel και Spafford (Garfinkel & G. Spafford, 1991) είναι η εξής λίστα:

- Το όνομά σας,
- Όνομα συζύγου,
- Όνομα πατέρα,
- Όνομα κατοικίδιου.,
- Όνομα του παιδιού σας,
- Ονόματα στενών σας φίλων και συνεργατών σας,
- Ονόματα αγαπημένων σας προσωπικοτήτων,
- Όνομα του εργοδότη σας,
- Οποιοδήποτε όνομα,
- Το όνομα που χρησιμοποιείτε στον υπολογιστή σας,
- Το κινητό σας τηλέφωνο,
- Ο αριθμός των πινακίδων σας,
- Οποιοσδήποτε αριθμός δημοσίου εγγράφου (ταυτότητα, ΑΦΜ, κτλ),
- Οποιαδήποτε ημερομηνία γέννησης.
- Άλλες πληροφορίες που μπορούν εύκολα να αντιστοιχηθούν με εσάς.

- Οποιοδήποτε όνομα χρήστη του υπολογιστή σας σε οποιαδήποτε μορφή.
- Οποιαδήποτε λέξη που υπάρχει στο αγγλικό, ελληνικό και οποιοδήποτε άλλο λεξικό.
- Κωδικός με επαναλαμβανόμενους χαρακτήρες,
- Πρότυπα απλά στο πληκτρολόγιο, όπως qwerty.
- Όλα τα παραπάνω γραμμένα ανάποδα
- Όλα τα παραπάνω προηγούμενα ή ακολουθούμενα από κάποιο ψηφίο.

4.4. ΚΟΙΝΟΙ ΚΑΙ ΑΝΘΕΚΤΙΚΟΙ ΚΩΔΙΚΟΙ ΠΡΟΣΒΑΣΗΣ

4.4.1. Οι 500 κοινοί κωδικοί πρόσβασης παγκοσμίως.

Όπως αποκάλυψε η ιστοσελίδα whatsmypass ('whatsmypass', 2008), από την πρώτη κιόλας στιγμή που άρχισαν οι άνθρωποι να χρησιμοποιούν τους κωδικούς πρόσβασης, σε σύντομο χρονικό διάστημα διαπιστώθηκε πως χρήστες έκαναν χρήση του ίδιου κωδικού πρόσβασης ολοένα και περισσότερο. Στην πραγματικότητα, οι άνθρωποι είναι τόσο προβλέψιμοι από επιτήδειους επιδρομείς (hackers) και έτσι κάνουν χρήση κάποιων κωδικών πρόσβασης οι οποίοι ανήκουν στην λίστα με τους πιο κοινούς κωδικούς. Η ιστοσελίδα αυτή παραθέτει αυτόν τον πίνακα και συμβουλεύει ότι αν τυχόν ο κωδικός πρόσβασης ενός χρήστη βρίσκεται σε αυτόν τον πίνακα, πρέπει να γίνει άμεσα αλλαγή αυτού καθώς έχει ήδη χρησιμοποιηθεί από χιλιάδες χρήστες στο παρελθόν και θεωρείται προβλέψιμος.
Ακολουθεί ο πίνακας των 500 πιο κοινών password με βάση την συγκεκριμένη ιστοσελίδα:

Πίνακας 4.1: Οι 500 κοινοί κωδικοί πρόσβασης παγκοσμίως.

NO	Top 1-100	Top 101–200	Top 201–300	Top 301–400	Top 401–500
1	123456	porsche	firebird	prince	rosebud
2	password	guitar	butter	beach	jaguar
3	12345678	chelsea	united	amateur	great
4	1234	black	turtle	7777777	cool
5	pussy	diamond	steelers	muffin	cooper
6	12345	nascar	tiffany	redsox	1313
7	dragon	jackson	zxcvbn	star	scorpio
8	qwerty	cameron	tomcat	testing	mountain
9	696969	654321	golf	shannon	madison
10	mustang	computer	bond007	murphy	987654
11	letmein	amanda	bear	frank	brazil
12	baseball	wizard	tiger	hannah	lauren
13	master	xxxxxxxx	doctor	dave	japan
14	michael	money	gateway	eagle1	naked
15	football	phoenix	gators	11111	squirt
16	shadow	mickey	angel	mother	stars
17	monkey	bailey	junior	nathan	apple
18	abc123	knight	thx1138	raiders	alexis
19	pass	iceman	porno	steve	aaaa
20	fuckme	tigers	badboy	forever	bonnie
21	6969	purple	debbie	angela	peaches
22	jordan	andrea	spider	viper	jasmine
23	harley	horny	melissa	ou812	kevin
24	ranger	dakota	booger	jake	matt
25	iwantu	aaaaaa	1212	lovers	qwertyui
26	jennifer	player	flyers	suckit	danielle
27	hunter	sunshine	fish	gregory	beaver
28	fuck	morgan	porn	buddy	4321

NO	Top 1-100	Top 101–200	Top 201–300	Top 301–400	Top 401–500
29	2000	starwars	matrix	whatever	4128
30	test	boomer	teens	young	runner
31	batman	cowboys	scooby	nicholas	swimming
32	trustno1	edward	jason	lucky	dolphin
33	thomas	charles	walter	helpme	gordon
34	tigger	girls	cumshot	jackie	casper
35	robert	booboo	boston	monica	stupid
36	access	coffee	braves	midnight	shit
37	love	xxxxxx	yankee	college	saturn
38	buster	bulldog	lover	baby	gemini
39	1234567	ncc1701	barney	cunt	apples
40	soccer	rabbit	victor	brian	august
41	hockey	peanut	tucker	mark	3333
42	killer	john	princess	startrek	canada
43	george	johnny	mercedes	sierra	blazer
44	sexy	gandalf	5150	leather	cumming
45	andrew	spanky	doggie	232323	hunting
46	charlie	winter	zzzzzz	4444	kitty
47	superman	brandy	gunner	beavis	rainbow
48	asshole	compaq	horney	bigcock	112233
49	fuckyou	carlos	bubba	happy	arthur
50	dallas	tennis	2112	sophie	cream
51	jessica	james	fred	ladies	calvin
52	panties	mike	johnson	naughty	shaved
53	pepper	brandon	xxxxx	giants	surfer
54	1111	fender	tits	booty	samson
55	austin	anthony	member	blonde	kelly
56	william	blowme	boobs	fucked	paul
57	daniel	ferrari	donald	golden	mine

NO	Top 1-100	Top 101–200	Top 201–300	Top 301–400	Top 401–500
58	golfer	cookie	bigdaddy	0	king
59	summer	chicken	bronco	fire	racing
60	heather	maverick	penis	sandra	5555
61	hammer	chicago	voyager	pookie	eagle
62	yankees	joseph	rangers	packers	hentai
63	joshua	diablo	birdie	einstein	newyork
64	maggie	sexsex	trouble	dolphins	little
65	biteme	hardcore	white	0	redwings
66	enter	666666	topgun	chevy	smith
67	ashley	willie	bigtits	winston	sticky
68	thunder	welcome	bitches	warrior	cocacola
69	cowboy	chris	green	sammy	animal
70	silver	panther	super	slut	broncos
71	richard	yamaha	qazwsx	8675309	private
72	fucker	justin	magic	zxcvbnm	skippy
73	orange	banana	lakers	nipples	marvin
74	merlin	driver	rachel	power	blondes
75	michelle	marine	slayer	victoria	enjoy
76	corvette	angels	scott	asdfgh	girl
77	bigdog	fishing	2222	vagina	apollo
78	cheese	david	asdf	toyota	parker
79	matthew	maddog	video	travis	qwert
80	121212	hooters	london	hotdog	time
81	patrick	wilson	7777	paris	sydney
82	martin	butthead	marlboro	rock	women
83	freedom	dennis	srinivas	xxxx	voodoo
84	ginger	fucking	internet	extreme	magnum
85	blowjob	captain	action	redskins	juice
86	nicole	bigdick	carter	erotic	abgrtyu

NO	Top 1-100	Top 101–200	Top 201–300	Top 301–400	Top 401–500
87	sparky	chester	jasper	dirty	777777
88	yellow	smokey	monster	ford	dreams
89	camaro	xavier	teresa	freddy	maxwell
90	secret	steven	jeremy	arsenal	music
91	dick	viking	11111111	access14	rush2112
92	falcon	snoopy	bill	wolf	russia
93	taylor	blue	crystal	nipple	scorpion
94	111111	eagles	peter	iloveyou	rebecca
95	131313	winner	pussies	alex	tester
96	123123	samantha	cock	florida	mistress
97	bitch	house	beer	eric	phantom
98	hello	miller	rocket	legend	billy
99	scooter	flower	theman	movie	6666
100	please	jack	oliver	success	albert

4.4.2. Μυστικότητα κωδικών πρόσβασης.

Σχετικά με την μυστικότητα των κωδικών πρόσβασης, συλλέχθηκαν πληροφορίες από μελέτη στο διαδίκτυο αναφορικά με τους τρόπους αποθήκευσης και τήρησης των κωδικών πρόσβασης των ηλεκτρονικών μέσων κρυφών. Επίσης ιδιαίτερα εποικοδομητική ήτα νη επικοινωνία και συζήτηση με χρήστες, ανεξαρτήτου ηλικίας, οι οποίοι παρέθεσαν τον τρόπο σκέψης τους και τις ανησυχίες τους από πιθανούς κινδύνους αποκάλυψης των κωδικών τους πρόσβασης. Έτσι λοιπόν, παραθέτονται οι παρακάτω προτάσεις/υποδείξεις προς τους χρήστες τους διαδικτύου για την εξασφάλιση της μυστικότητας των κωδικών τους:

1. **Μην τους αποκαλύπτετε σε άλλους.** Να κρατάτε τους κωδικούς πρόσβασής σας μυστικούς από τους φίλους και τα μέλη της οικογένειάς σας που μπορεί να τους αποκαλύψουν σε άλλα, μη έμπιστα πρόσωπα. Ιδιαίτερη προσοχή από τα μέλη της οικογενείας πρέπει να δοθεί στα

παιδιά. Κι αυτό διότι δεν έχουν την αίσθηση του κινδύνου μιας τέτοιας αποκάλυψης. Ενδέχεται να το αποκαλύψουν και σε άτομα που πιθανόν να μην γνωρίζουμε ότι το έχουν πει με συνέπεια να μην αντιληφθούμε την ανάγκη άμεσης αλλαγής του. Μόνες εξαιρέσεις είναι οι κωδικοί πρόσβασης που πρέπει να μοιράζεστε με άλλους, όπως είναι ο κωδικός πρόσβασης για το διαδικτυακό τραπεζικό σας λογαριασμό που πρέπει να είναι κοινός με τη σύζυγό σας.

2. **Να προστατεύετε τους καταγεγραμμένους κωδικούς πρόσβασης.** Προσέχετε που αποθηκεύετε τους κωδικούς πρόσβασης που καταγράφετε ή σημειώνετε. Μην αφήνετε τα αρχεία σας έκθετα, όπως δεν θα αφήνατε και τα στοιχεία που προστατεύουν.
3. **Ποτέ να μην αποκαλύπτετε τον κωδικό πρόσβασης μέσω ηλεκτρονικού ταχυδρομείου και να μην τον καταγράφετε σε αιτήσεις που αποστέλλονται μέσω ηλεκτρονικού ταχυδρομείου.** Οποιοδήποτε μήνυμα ηλεκτρονικού ταχυδρομείου που ζητά τον κωδικό πρόσβασης ή σας ζητά να μεταβείτε σε κάποια τοποθεσία του Διαδικτύου για να επιβεβαιώσετε τον κωδικό σας είναι απάτη. Μια τράπεζα δεν θα ζητήσει ποτέ να επιβεβαιώσουμε τον κωδικό πρόσβασής μας στέλνοντας με ηλεκτρονικό ταχυδρομείο. Αυτό αφορά και τις αιτήσεις αξιόπιστων εταιρειών ή προσώπων. Τα μηνύματα ηλεκτρονικού ταχυδρομείου μπορούν να υποκλαπούν κατά τη μεταφορά και τα μηνύματα που απαιτούν την καταγραφή πληροφοριών ίσως να μην προέρχονται από τον αποστολέα που ισχυρίζονται. Οι απάτες ηλεκτρονικού "ψαρέματος" (phishing) μέσω Internet χρησιμοποιούν μηνύματα ηλεκτρονικού ταχυδρομείου για να σας παρασύρουν να αποκαλύψετε ονόματα χρήστη και κωδικούς πρόσβασης, να κλέψουν στοιχεία ταυτότητας και άλλα.
4. **Να αλλάζετε τακτικά τους κωδικούς πρόσβασης.** Έτσι μπορεί να κρατήσετε μακριά τους εγκληματίες και τους άλλους κακόβουλους χρήστες. Η ισχύς του κωδικού πρόσβασης θα σας βοηθήσει να τον διατηρήσετε για περισσότερο χρόνο. Ένας κωδικός πρόσβασης μικρότερος από 8 χαρακτήρες θα παραμένει ισχυρός για μία περίπου εβδομάδα, ενώ ένας κωδικός πρόσβασης 14 χαρακτήρων ή μεγαλύτερος (που ακολουθεί και τους άλλους κανόνες που περιγράφηκαν παραπάνω) μπορεί να παραμείνει ισχυρός για χρόνια. Το ίδιο υποστηρίζει και η επίσημη ιστοσελίδα της

microsoft (‘Microsoft’, 2007) η οποία επιπροσθέτως τονίζει ότι δεν είναι μόνο οι κακόβουλοι χρήστες από τους οποίους κινδυνεύει ο κωδικός μας αλλά και το "ηλεκτρονικό ψάρεμα" (phishing) ή το "keyllogging". Επειδή όμως το κόστος της εκπαίδευσης για την αντιμετώπιση του ηλεκτρονικού ψαρέματος είναι μεγάλο, τότε η εύκολη λύση είναι να καταφύγουμε στην συχνή αλλαγή του password μας.

5. **Μην πληκτρολογείτε κωδικούς πρόσβασης σε υπολογιστές που δεν ελέγχετε.** Οι υπολογιστές που υπάρχουν σε Ίντερνετ καφέ, εργαστήρια υπολογιστών, συστήματα κοινής χρήσης, συστήματα σε κιόσκια και αίθουσες αναμονής αεροδρομίων δεν θα πρέπει να θεωρούνται ασφαλείς για οποιαδήποτε χρήση εκτός από ανώνυμη περιήγηση στο Διαδίκτυο. Να μην χρησιμοποιείτε τους υπολογιστές αυτούς για να ελέγχετε το ηλεκτρονικό σας ταχυδρομείο, να μπαίνετε σε χώρους ηλεκτρονικής συνομιλίας, να ελέγχετε το τραπεζικό σας υπόλοιπο και το επιχειρηματικό σας ταχυδρομείο ή για να μπαίνετε σε οποιονδήποτε άλλον λογαριασμό που να απαιτεί όνομα χρήστη και κωδικό πρόσβασης. Οι εγκληματίες μπορούν να αγοράσουν πολύ φθηνά συσκευές σύνδεσης με το πληκτρολόγιο, που εγκαθίστανται σε ελάχιστο χρόνο. Οι συσκευές αυτές επιτρέπουν στους κακόβουλους χρήστες να συλλέγουν μέσω Internet όλες τις πληροφορίες που πληκτρολογούνται σε έναν υπολογιστή.

4.5. Επιλογή κενού κωδικού πρόσβασης.

Ένας κενός κωδικός πρόσβασης (κανένας κωδικός πρόσβασης) στο λογαριασμό σας είναι ασφαλέστερος από έναν ανίσχυρο κωδικό πρόσβασης, όπως ο "1234". Οι εισβολείς μπορούν να μαντέψουν εύκολα κάποιον απλούστερο κωδικό πρόσβασης αλλά, όταν υπάρχουν υπολογιστές που χρησιμοποιούν Windows XP, αυτοί δεν μπορούν να προσπελαστούν από απόσταση αν ο λογαριασμός χρήστη δεν έχει κωδικό πρόσβασης (αυτή η επιλογή δεν είναι διαθέσιμη για λειτουργικό σύστημα Microsoft Windows 2000, Windows Me ή νεότερες εκδόσεις). Μπορείτε να επιλέξετε να χρησιμοποιήσετε κενό κωδικό πρόσβασης στο λογαριασμό του υπολογιστή σας εάν πληρούνται τα παρακάτω κριτήρια:

1. Έχετε μόνον έναν υπολογιστή ή έχετε πολλαπλούς υπολογιστές αλλά δεν χρειάζεται να προσπελαύνετε πληροφορίες από τον έναν υπολογιστή στον άλλον
2. Ο υπολογιστής είναι φυσικά ασφαλής (εμπιστεύεστε όλους όσους έχουν φυσική πρόσβαση στον υπολογιστή).

4.6. Χρήση διαφορετικών κωδικών πρόσβασης.

Η ιστοσελίδα PasswordResearch (‘PasswordResearch’, 2011) αναφέρει ότι σύμφωνα με μια μελέτη σχετικά με τη "Θέση και συμπεριφορά απέναντι στην χρήση των κωδικών πρόσβασης στο διαδίκτυο" το 70% των ανθρώπων χρησιμοποιούν διαφορετικούς κωδικούς πρόσβασης για διαφορετικές ιστοσελίδες.
Η μελέτη έγινε σε ένα δείγμα εκατόν εικοσιτεσσάρων ατόμων (124) και παρακάτω αναφέρονται οι ερωτήσεις οι οποίες υποβλήθηκαν σε αυτό το δείγμα και ταυτόχρονα εμφανίζεται και το στατιστικό ποσοστό των απαντήσεών τους.

- Διαθέτετε ένα κοινό κωδικό πρόσβασης για κάθε ιστοσελίδα που απαιτεί είσοδο σε αυτήν;

Πίνακας 4.2: Κοινό password για κάθε ιστοσελίδα.

Ναι	30,4%
Όχι	69,6%

- Πόσες από τις ιστοσελίδες που επισκέπτεστε σας ζητούν κωδικό πρόσβασης;

Πίνακας 4.3: Ιστοσελίδες που ζητούν password.

Καμία	1,6%
Μία	10,4%
Δύο με πέντε	57,6%
Έξι με δέκα	18,4%
Περισσότερες από δέκα	11,2%

- Αν επισκέπτεστε συχνά ιστοσελίδες οι οποίες ζητάνε κάποιο κωδικό πρόσβασης, σε πόσες από αυτές χρησιμοποιείτε ένα κοινό κωδικό;

Πίνακας 4.4: Κοινός κωδικός σε όλα τα site.

Ένα	24%
Δύο με πέντε Έξι με δέκα	64,8% 6,4%
Περισσότερες από δέκα	1,6%

Σε έρευνα που διεξήχθη στα πλαίσια της παρούσας εργασίας *(βλ. κεφάλαιο 5)* ***από το δείγμα των 594 χρηστών, ένα ποσοστό του 32% απάντησε ότι χρησιμοποιεί ένα και μόνο κωδικό για τις διάφορες εφαρμογές που απαιτούν είσοδο του χρήστη.***

4.7. Συνέπειες αδύναμων κωδικών πρόσβασης.

Αν κάποιος τελικά καταφέρει να αποκτήσει τον κωδικό πρόσβασης ενός χρήστη, τότε πιθανόν να αρχίσει να χρησιμοποιεί τον λογαριασμό του για να δει προσωπικά μας δεδομένα, όπως για παράδειγμα το email μας, τραπεζικούς λογαριασμούς, μηνύματα τηλεφώνου. Τότε πιθανό είναι το άτομο αυτό να προσπαθήσει να αλλάξει, να μετατρέψει ή ακόμη και να καταστρέψει αρχεία ή ακόμη σε χειρότερη περίπτωση να αναλάβει τον πλήρη έλεγχο του υπολογιστή του χρήστη. Αυτό σαφώς συνεπάγεται ότι θα έχει και την δυνατότητα για να προβεί σε παράνομες δραστηριότητες εκ μέρους του-σε αρκετές περιπτώσεις είναι δύσκολο να ανακαλυφθεί ο πραγματικός ένοχος και τότε προφανώς θα θεωρείται ως ένοχος ο απλός χρήστης.

4.8. ΄ΕΡΕΥΝΕΣ

4.8.1. Έρευνα του πανεπιστημίου του Cambridge.

Το Σεπτέμβριο του 2000 ανακοινώθηκαν από στο πανεπιστήμιο του Cambridge έρευνες σχετικά με την «υπενθύμιση και ασφάλεια των κωδικών» (Yan et.al., 2000; 'Password Memorability and Security', 2004). Σε κάποια από αυτές, που διεξήχθη τον Οκτώβριο του 1999, έγινε ένα πείραμα στο οποίο συμμετείχαν 288 φοιτητές του πανεπιστημίου, με θέμα την επιλογή των κωδικών πρόσβασης.

Μέθοδος:

Στο συγκεκριμένο πείραμα, οι φοιτητές χωρίστηκαν με βάση την επιλογή τους σε τρεις ομάδες.

- *Την ομάδα ελέγχου*, στην οποία δίνονταν η συμβουλή να δημιουργήσουν password με το λιγότερο εφτά (7) χαρακτήρες εκ των οποίων τουλάχιστον ένα να μην είναι γράμμα.
- *Την ομάδα τυχαίου κωδικού πρόσβασης*, οι οποίοι αποτελούνται από όλα τα κεφαλαία γράμματα της αλφαβήτας (A-Z) και όλα τα νούμερα (1-9 με μήκος οχτώ (8) χαρακτήρων.
- *Την ομάδα φράσης*, στην οποία οι φοιτητές διάλεξαν ως password μία φράση που θυμούνται.

Οι μελετητές της έρευνας πίστευαν ότι η *ομάδα τυχαίου κωδικού πρόσβασης* θα είχε πιο ανθεκτικά password από την *ομάδα φράσης*, αλλά θα ήταν δυσκολότερο γι' αυτούς να τα θυμούνται ή ότι θα τα ξεχνούσαν εύκολα. Ωστόσο, ένα μήνα μετά από την δημιουργία των password εξετάστηκαν τρείς τύποι επιθέσεων που έγιναν σε αυτά.

1. Λεξιλογικές επιθέσεις.
2. Επιθέσεις βασισμένες στις πληροφορίες του χρήστη.
3. Άμεση επίθεση.

Πίνακας 4.5: Συμμετέχοντες σε κάθε ομάδα.

Ομάδα ελέγχου	95
Ομάδα τυχαίου password	96
Ομάδα φράσης	97

Συμπεράσματα της έρευνας:

1. Ο μέσος όρος χαρακτήρων που χρησιμοποίησαν ήταν 7,6 χαρακτήρες και είχαν ελάχιστα μεγαλύτερο μήκος από άλλους φοιτητές που δεν συμμετείχαν στο πείραμα και δεν είχαν διδαχθεί τις ανάλογες συμβουλές.
2. Η πιο επιτυχής μέθοδος επιθέσεων ήταν η λεξιλογική.
3. Σε καμία περίπτωση δεν ήταν επιτυχής αυτή που βασίζεται στις πληροφορίες του χρήστη για διάφορους λόγους. Προφανώς επειδή δεν υπήρχε αρκετό στοιχείο τους χρήστες βασισμένο σε προσωπικές τους πληροφορίες.
4. Όλοι οι κωδικοί πρόσβασης που είχαν μήκος έξι χαρακτήρων παραβιάστηκαν. Παρά τις όποιες συμβουλές των ερευνητών, οι χρήστες τις αγνόησαν και διάλεξαν μικρό αριθμό χαρακτήρων για τον κωδικό τους. Συνεπώς ήταν αδύναμοι.
5. Όπως αναφέρθηκε παραπάνω όλοι οι κωδικοί με μήκος μικρότερο των έξι χαρακτήρων παραβιάστηκαν. Πρέπει να σημειωθεί όμως ότι οι περισσότεροι ανήκουν στην ομάδα ελέγχου και στην ομάδα τυχαίου κωδικού. Αυτό φυσικά αντικρούει τις αρχικές προβλέψεις των ερευνητών.
6. Οι κωδικοί οι οποίοι παραβιάστηκαν από την ομάδα τυχαίου κωδικού και την ομάδα φράσης, οφείλονται σε λέξεις που υπάρχουν σε λεξικά, σε επαναλαμβανόμενες λέξεις και αριθμούς.

Αυτά τα αποτελέσματα, μαζί με το γεγονός ότι πολλοί χρήστες διάλεξαν μήκος έξι χαρακτήρων για τον κωδικό τους, μας παρέχουν ένα αξιόπιστο συμπέρασμα: **Οι χρήστες δεν συμμορφώνονται με τις γενικές συμβουλές για την κατασκευή των κωδικών πρόσβασης.** Επίσης παρατηρήθηκε ότι κανείς δεν χρησιμοποίησε ειδικούς χαρακτήρες, εκτός από την *ομάδα φράσης*. Άρα με βάση και αυτή την έρευνα, για να έχει κανείς ανθεκτικό κωδικό πρόσβασης προτείνεται να κάνει χρήση συνδυασμών σε μεγάλο μήκος χαρακτήρων γραμμάτων, αριθμών και ειδικών συμβόλων.

7. Τέλος, παρατηρήθηκε ότι ελάχιστοι χρήστες ζήτησαν να αλλαχτεί ο κωδικός τους. Συγκεκριμένα, μετά από ένα διάστημα τριών μηνών, ένα ποσοστό της τάξης του 2% ζήτησε την αλλαγή του κωδικού τους. Στον παρακάτω πίνακα παρουσιάζουμε ανά ομάδα τις αλλαγές των κωδικών.

4.8.2. Έρευνες Usability news.

Σε μία άλλη έρευνα που διεξήχθη το 2006 από τον Shannon Riley (Riley, 2006) δημοσιεύτηκαν πολύτιμα συμπεράσματα σχετικά με:

1. τους τύπους και τον αριθμό των διαφορετικών κωδικών πρόσβασης που χρησιμοποιούν οι χρήστες
2. πρακτικές δημιουργίας και αποθήκευσης των κωδικών
3. πρακτικές που θα έπρεπε να χρησιμοποιούνται για την αποθήκευση των κωδικών
4. γενικές δημογραφικές πληροφορίες.

Στην έρευνα συμμετείχαν 315 άτομα ηλικίας από 18 έως 58, με μέσο όρο ηλικίας συμμετεχόντων 25,34.
Συγκεκριμένα αναφέροντας τα γενικά χαρακτηριστικά των κωδικών η έρευνα έδειξε ότι το 35% (112) των συμμετεχόντων χρησιμοποιούν ένα προκαθορισμένο αριθμό χαρακτήρων στα password τους.

Διάγραμμα 4.1: Ποσοστό προκαθορισμένου αριθμού χαρακτήρων σε κωδικό.

Από αυτούς τους χρήστες βρέθηκε ότι κατά μέσο όρο χρησιμοποιούν 6,84 χαρακτήρες, αριθμός που θεωρείται μικρός για την ασφάλειά τους. Στον παρακάτω πίνακα παρουσιάζονται αποτελέσματα τις έρευνας για την συχνότητα με τη οποία οι χρήστες αλλάζουν τους κωδικούς τους, με την προϋπόθεση ότι αυτό δεν απαιτείται από το σύστημα.

Πίνακας 4.6: Συχνότητα αλλαγής κωδικού.

Συχνότητα	Απαντήσεις	Ποσοστό
Ποτέ	166	52.70%
Κάθε χρόνο	53	16.83%
Κάθε έξι μήνες	44	13.97%
Κάθε τρείς μήνες	38	12.06%
Κάθε μήνα	12	3.81%
Κάθε εβδομάδα	1	0.32%
Άγνωστο	1	0.32%
Σύνολο	315	100.00%

Επίσης στη έρευνα αυτή εξετάστηκαν οι τρόποι δημιουργίας των password κατά την οποία αποδείχθηκε ότι το 85.7% (270) των συμμετεχόντων χρησιμοποιούν μικρά γράμματα και το 56.5% (178) κάνουν χρήση αριθμών. Αναλυτικά στον παρακάτω πίνακα παρουσιάζονται τα ποσοστά των χρήσεων διαφόρων τακτικών για την δημιουργία κωδικού πρόσβασης, ενώ παρατίθεται και το αντίστοιχο σχήμα.

Πίνακας 4.7: Συχνότητα πρακτικών δημιουργίας κωδικού πρόσβασης.

Πρακτικές	Ποτέ	Πολύ Σπάνια	Σπάνια	Περιστασια-κά	Πολύ Συχνά	Πάντα
Κεφαλαία γράμματα	120	101	35	32	21	5
Αριθμοί	17	23	26	71	107	71
Ειδικοί χαρακτήρες	222	36	23	22	6	6
Κενά	225	31	21	5	2	1
Ονόματα φίλων, συγγενών κτλ	144	30	23	50	46	22
Προσωπικές πληροφορίες (ημ/νια γέννησης, αρ.τηλεφώνου)	48	11	17	66	109	64
Γεωγραφικές τοποθεσίες	215	33	21	20	19	6
Απλές ακολουθίες χαρακτήρων (π.χ1234)	225	36	23	28	11	4

Πρακτικές	Ποτέ	Πολύ Σπάνια	Σπάνια	Περιστασια-κά	Πολύ Συχνά	Πάντα
Επαναλαμβανόμενοι χαρακτήρες	225	27	13	15	5	0

Διάγραμμα 4.2: Τρόποι δημιουργίας κωδικών.

Αναφορικά με τους τρόπους αποθήκευσης των password, μέσα από την έρευνα αποδείχθηκε ότι το 54.6% (177) των συμμετεχόντων χρησιμοποιούν τον ίδιο κωδικό για διάφορους λογαριασμούς ενώ το 33% (104) χρησιμοποιεί μια μικρή ποικιλία κωδικών όπως φαίνεται και στο ακόλουθο σχήμα:

Διάγραμμα 4.3: Αριθμός κωδικών από τους χρήστες.

Στον παρακάτω πίνακα παρουσιάζονται τα ποσοστά χρήσης των διαφόρων πρακτικών αποθήκευσης των κωδικών πρόσβασης.

Πίνακας 4.8: Συχνότητα πρακτικών αποθήκευσης κωδικών.

Πρακτικές	Ποτέ	Πολύ Σπάνια	Σπάνια	Περιστασιακά	Πολύ Συχνά	Πάντα
Λίστα από password	130	66	27	44	20	27
Ίδια password	26	18	20	74	134	43
Διαφορετικά password	69	25	36	81	85	19
Απομνημόνευση	94	33	31	67	60	30
Χειρόγραφα	294	8	3	6	3	1

Τέλος, στα πλαίσια της έρευνας του Riley, εξετάστηκε η αντίληψη των χρηστών απέναντι στην αναγνώριση πρακτικών για τρόπους δημιουργίας και χρήσης των κωδικών. Διαπιστώθηκε λοιπόν πως ένα ποσοστό της τάξης του 50.8% (106) ήταν σε θέση να αναγνωρίσει μια πρακτική δημιουργίας ενός password, ενώ το 62.9% (198) μπορούσαν να διακρίνουν και να αναγνωρίσουν ποιοί κωδικοί είναι ασφαλείς.

4.9. ΕΠΙΛΟΓΟΣ

Σε αυτό το κεφάλαιο έγινε μια αναφορά στους κωδικούς πρόσβασης. Αναφέρθηκε η αναγκαιότητά τους και ο βαθμός στον οποίο η μυστικότητά τους αποτελεί ασφάλεια για τους κοινούς χρήστες.

Παρατηρήθηκε μέσα από μελέτες ότι οι περισσότεροι χρήστες αν και έχουν ενημερωθεί ή γνωρίζουν το πόσο σημαντικό είναι να διατηρούνται μυστικοί οι κωδικοί πρόσβασης και γενικότερα να είναι ανθεκτικοί, αψηφούν τις όποιες συμβουλές.

Συμπεραίνεται ότι προτιμούν μικρούς και εύκολους κωδικούς, έτσι ώστε να διευκολύνονται και να τους θυμούνται. Γίνεται χρήση σημείων στίξης και διάφορων συμβόλων, ωστόσο η συχνότητά τους και η κατανομή γραμμάτων και αριθμών δεν εμφανίζεται ικανοποιητική σε βαθμό που να αυξάνει την πολυπλοκότητά τους.

Σε αρκετές περιπτώσεις οι χρήστες χρησιμοποιούν προσωπικά δεδομένα ή μικρό μήκος χαρακτήρων με αποτέλεσμα οι κωδικοί τους να είναι αδύναμοι και να κινδυνεύει η ασφάλειά τους.

Περαιτέρω έρευνα διεξήχθη στα πλαίσια αυτής της πτυχιακής για τους κωδικούς πρόσβασης που χρησιμοποιούν οι Έλληνες χρήστες στα ηλεκτρονικά μέσα, τα συμπεράσματα της οποία αναλύονται στο επόμενο κεφάλαιο.

ΚΕΦΑΛΑΙΟ 5

5. Η ΚΑΤΑΣΤΑΣΗ ΣΤΗΝ ΕΛΛΑΔΑ

5.1. ΠΡΟΛΟΓΟΣ

Σε αυτό το κεφάλαιο θα γίνει αναφορά στους κωδικούς πρόσβασης που χρησιμοποιούνται από τους Έλληνες χρήστες. Η επισήμανση αυτή θα γίνει μέσα από μελέτη που διεξήχθη στα πλαίσια της εργασίας. Μέσω της έρευνας θα γίνει εκτενής ανάλυσή τους σχετικά με τρόπο επιλογής τους κατά την δημιουργία τους αλλά και την κατανομή των γραμμάτων που παρουσιάζουν. Επίσης ένα άλλο σημαντικό αντικείμενο που θα εξεταστεί όσον αφορά τους κωδικούς πρόσβασης, είναι η ικανότητα των Ελλήνων χρηστών να αναγνωρίσουν ισχυρούς κωδικούς πρόσβασης και να τους διακρίνουν από τους αδύναμους. Αντίστοιχη έρευνα δεν έχει διεξαχθεί στο παρελθόν και **για πρώτη φορά στην Ελλάδα θα παρουσιαστούν οι αδύναμοι κωδικοί πρόσβασης που θα πρέπει να αποφεύγουν οι Έλληνες χρήστες.**

5.2. Θέμα της έρευνας.

Το θέμα της έρευνας αφορά τους κωδικούς πρόσβασης από Έλληνες χρήστες. Ενδεικτικά αναφέρεται ότι μελετήθηκαν τρόποι και μέθοδοι δημιουργίας των κωδικών από τους χρήστες, επιλογές χαρακτήρων για την σύνθεση ενός κωδικού, διάφορες πρακτικές απομνημόνευσης και αποθήκευσής τους, αλλά και η ικανότητα των χρηστών να διακρίνουν ισχυρούς κωδικούς. **Σε αυτό το σημείο κρίνεται σκόπιμο να αναφερθεί ότι μέχρι σήμερα δεν έχει διεξαχθεί αντίστοιχη έρευνα ή μελέτη στην Ελλάδα σχετικά με την πολυπλοκότητα και την χρήση των κωδικών πρόσβασης.**

5.3. Τρόπος διεξαγωγής της έρευνας.

Η έρευνα διεξήχθη μέσω ερωτηματολογίου. Η επιλογή αυτή έγινε καθώς θεωρήθηκε ότι με αυτό τον τρόπο μπορούσαν να προσεγγιστούν περισσότεροι χρήστες. Μέσω του διαδικτύου και των διαφόρων μέσων επικοινωνίας που

υπάρχουν σε αυτό (π.χ e-mail, facebook, forum) υπάρχει η δυνατότητα αποστολής ερωτηματολογίου σε μεγάλο πλήθος ανθρώπων, γνωστών και φίλων, καθώς και η προώθησή του σε περαιτέρω αριθμό χρηστών από τους τελευταίους. Επίσης επιλέχθηκε αυτή η μέθοδος συμπλήρωσης του ερωτηματολογίου, καθώς παρέχεται η δυνατότητα να απαντηθεί από τους χρήστες οποιαδήποτε στιγμή, ενώ πρέπει να αναφέρουμε πως το ερωτηματολόγιο ήταν ενεργό για ένα χρονικό διάστημα περίπου 5 μηνών.

5.4. Περιεχόμενο ερωτηματολογίου.

Το ερωτηματολόγιο απαρτίζεται από ερωτήσεις οι οποίες επιλέχθηκαν με κριτήριο την συμβολή τους στην εξαγωγή στατιστικών συμπερασμάτων, γεγονός που είναι ιδιαίτερα ενδιαφέρον αφού ανάλογες προσπάθειες δεν υπάρχουν καταγεγραμμένες για την Ελλάδα. Βεβαίως κάτι τέτοιο μπορεί και πρέπει να αποτελέσει την βάση για περαιτέρω έρευνα. Συνολικά το ερωτηματολόγιο αποτελείται από δεκατρείς (13) ερωτήσεις, αριθμός τέτοιος ώστε η έκτασή του να μην κουράσει τους συμμετέχοντες. Οι ερωτήσεις επιλέχθηκαν από μια πληθώρα διαθέσιμων επιλογών και έπειτα από κατάλληλη μελέτη και συζήτηση με ειδικούς αλλά και χρήστες επιλέχθηκαν οι θεωρούμενες πιο κατάλληλες.

Μέθοδος

Συμμετέχοντες:

Στην έρευνα συμμετείχαν 594 άτομα, εκ των οποίων αρκετοί έχουν άμεση σχέση με το αντικείμενο της πληροφορικής (πτυχιούχοι, φοιτητές). Το ερωτηματολόγιο αναρτήθηκε στην ιστοσελίδα φοιτητών του τμήματος Πληροφορικής το λεγόμενο «steki» (http://steki.it.teithe.gr/) και προωθήθηκε σε επαφές μέσω του διαδικτύου (facebook, msn).

Το ερωτηματολόγιο τοποθετήθηκε στο διαδίκτυο, είναι ακόμη σε ισχύ και βρίσκεται στον ακόλουθο σύνδεσμο:

http://tinyurl.com/pass-questions

Τα στοιχεία των συμμετεχόντων είναι ανώνυμα, γεγονός που σημαίνει ότι οι συμμετέχοντες απάντησαν ειλικρινά χωρίς ενδοιασμούς και συνεπώς δεν υπάρχει αλλοίωση των αληθινών τους απόψεων και πρακτικών που ακολουθούν.

Συμπερασματικά θεωρείται ότι η έρευνα παρέχει αληθή αποτελέσματα τα οποία θα αναλυθούν και θα σχολιαστούν παρακάτω.

Σχετικά με τον τρόπο λήψης των απαντήσεων, το ερωτηματολόγιο έπειτα από την αποστολή του στο προαναφερόμενο δείγμα, μετά από κάθε μία συμπλήρωσή του, επιστρεφόταν ανώνυμα για την ανάλυση και εξαγωγή στατιστικών αποτελεσμάτων. Η έρευνα διεξήχθη από τις 8 Μαρτίου 2011 έως 30 Σεπτεμβρίου 2011.

Ωστόσο είναι πιστευτό ότι ο αριθμός των συμμετεχόντων θα μπορούσε να είναι και μεγαλύτερος. Λόγω της καθολικής διείσδυσης του διαδικτύου και των εφαρμογών του (msn, facebook, κτλ.) θα μπορούσε μέσω των προωθήσεων σε επαφές φίλων και γνωστών να συγκεντρωθεί ένας αρκετά μεγαλύτερος αριθμός δείγματος. Παρόλα αυτά κάτι τέτοιο δεν συνέβη είτε γιατί κάποιοι το θεώρησαν επιπόλαιο ή όχι τόσο σημαντικό να συμμετάσχουν στην έρευνα, είτε γιατί θεώρησαν ότι κινδυνεύουν κατά κάποιο τρόπο να φανερώσουν τον τρόπο σκέψης τους για την κατασκευή του κωδικού τους. Αυτό μάλιστα ενισχύει το γεγονός ότι εφόσον δεν έχουν διεξαχθεί αντίστοιχες έρευνες, είναι πιθανόν κάποιοι χρήστες να φοβήθηκαν να συμμετάσχουν στην έρευνα. Αναμφισβήτητα ένας μεγαλύτερος αριθμός δείγματος θα έδινε πιο σαφή συμπεράσματα εκπροσωπώντας μεγαλύτερο πλήθος χρηστών.

5.5. Αποτελέσματα-Συμπεράσματα

Στην παράγραφο αυτή αναφέρονται τα συμπεράσματα και αποτελέσματα που εξάγονται από την έρευνα. Αναλυτικά οι ερωτήσεις του ερωτηματολογίου σε συνδυασμό με όλα τα στατιστικά του αποτελέσματα παρουσιάζονται στο παράρτημα Δ'.

Σε αυτήν την φάση θα αναδειχτούν τα κυριότερα σημεία που κρίνεται απαραίτητο να τονιστούν από την εξαγωγή αποτελεσμάτων της έρευνας. Οι ερωτήσεις που τέθηκαν δεν ήταν τυχαίες, αλλά έγινε προσπάθεια όσο το δυνατόν καλύτερα να εκτεθούν ζητήματα της καθημερινότητας που αφορούν τους χρήστες κυρίως του διαδικτύου. Τέτοια ζητήματα όπως θα σχολιαστούν παρακάτω είναι τα εξής:

- ✓ οι κωδικοί που έχουν χρησιμοποιηθεί στο παρελθόν,
- ✓ η συνήθεια αλλαγής του κωδικού πρόσβασης,
- ✓ η πολυπλοκότητα του κωδικού που χρησιμοποιούν,

- ✓ η ικανότητα διάκρισης αδύναμων κωδικών πρόσβασης,
- ✓ ο αριθμός κωδικών που χρησιμοποιεί συνολικά ένας χρήστης,
- ✓ τρόποι απομνημόνευσης ενός κωδικού,
- ✓ και η πιθανότητα αποκάλυψής του.

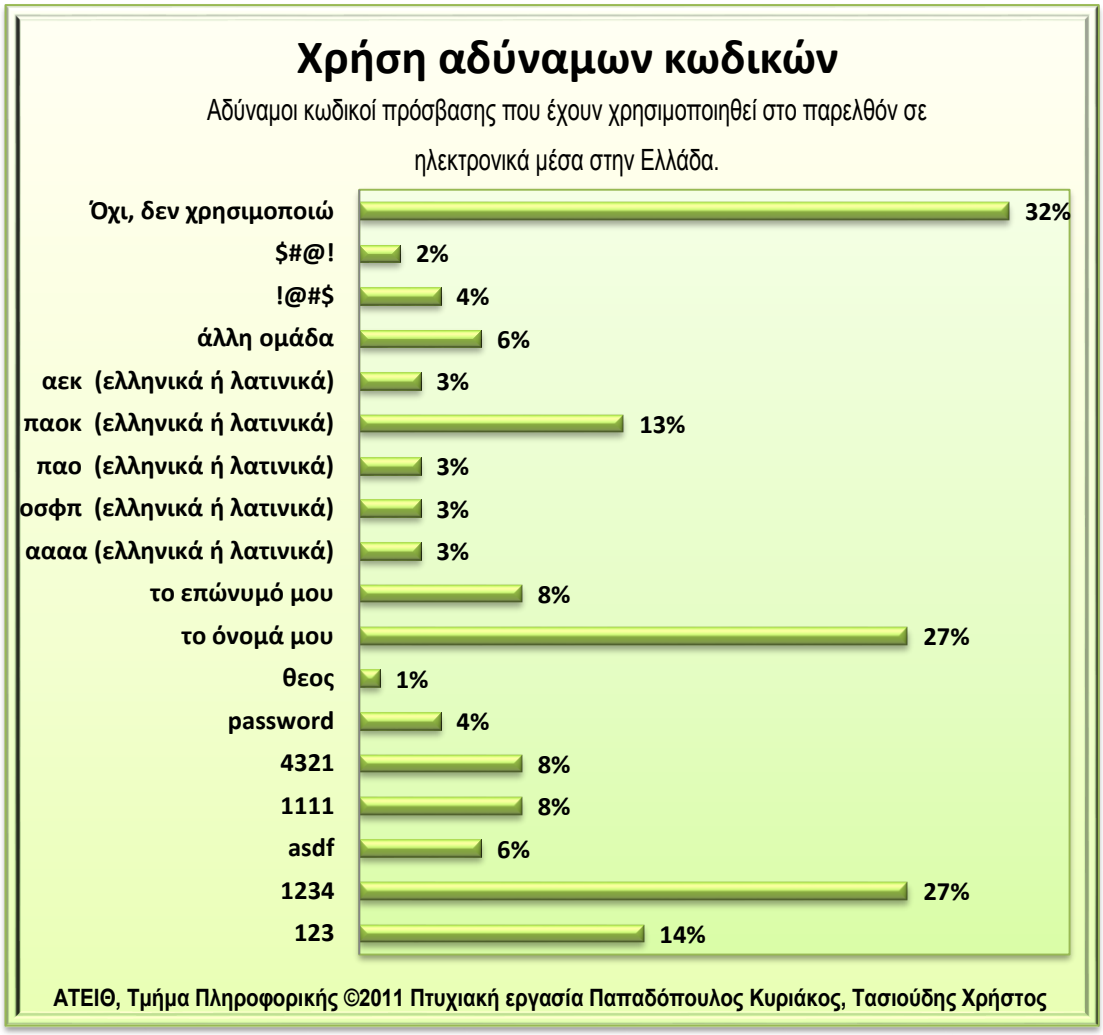
5.5.1. Παλαιότεροι συνηθισμένοι κωδικοί.

Στα πλαίσια της μελέτης ζητήθηκε από το δείγμα να αποκαλυφθούν κωδικοί που έχουν χρησιμοποιήσει στο παρελθόν και αντικαταστάθηκαν για λόγους ασφαλείας. Έτσι, παρουσιάστηκε μία λίστα με ενδεχόμενους αδύναμους κωδικούς, όπου οι χρήστες είχαν την δυνατότητα να επιλέξουν περισσότερους από κανέναν, έως περισσότερους από έναν.

Ένα μεγάλο ποσοστό του 32% απάντησε πως δεν έχει χρησιμοποιήσει κανέναν από αυτούς τους κωδικούς στο παρελθόν. Από τις πιο υψηλές επιλογές χρήσης όμως ήταν ο κωδικός "1234" με 27% όπως και το όνομα του χρήστη με το ίδιο ποσοστό, ενώ ακολουθεί με 14% το "123" και με 13% η αθλητική ομάδα "παοκ".

Εδώ είναι προφανές ότι η εντοπιότητα των ερευνητών και των γνωστών τους που δέχθηκαν να απαντήσουν επηρεάζει το δείγμα (bias). Προφανώς σε περιοχές τις Νότιας Ελλάδας ένας τέτοιος κωδικός πρόσβασης θα αντικατασταθεί από δημοφιλείς στην περιοχή αθλητικές ομάδες (Ολυμπιακός, ΟΣΦΠ, ΠΑΟ,ΑΕΚ κτλ.) Το εύρημα παραμένει ως η αγαπημένη ομάδα του ερωτηθέντος.

Στο παρακάτω διάγραμμα παρουσιάζονται τα ποσοστά των επιλογών όλων των πιθανών αδύναμων κωδικών που έχουν χρησιμοποιηθεί στο παρελθόν.



Διάγραμμα 5.1: Αδύναμοι κωδικοί πρόσβασης που έχουν χρησιμοποιηθεί στο παρελθόν σε ηλεκτρονικά μέσα στην Ελλάδα.

Σημείωση: Από τους παραπάνω κωδικούς που βρίσκονται στο διάγραμμα 5.1 υπάρχουν κάποιοι οι οποίοι είναι ακολουθιακοί. Κωδικοί όπως "$#@!" και "!@#$" αποτελούνται από χαρακτήρες που βρίσκονται σε διαδοχική σειρά στο πληκτρολόγιο με την χρήση του πλήκτρου Shift.

5.5.2. Συνήθεια αλλαγής κωδικού πρόσβασης.

Όπως έχει αναφερθεί σε προηγούμενο κεφάλαιο, προτείνεται γενικά στους χρήστες του διαδικτύου να αλλάζουν τακτικά τον κωδικό τους πρόσβασης. Ένας μικρός κωδικός, με πλήθος χαρακτήρων λιγότερο από οχτώ, μπορεί να θεωρηθεί ασφαλής μόνο για μικρό χρονικό διάστημα (Yan et.al., 2000).

Από το σύνολο των χρηστών που συμμετείχαν στην έρευνα, διαπιστώθηκε πως μόλις ένα μικρό ποσοστό του 13% προχωράει σε συχνές αλλαγές των κωδικών τους, ενώ το 39% των χρηστών δεν έχει αλλάξει ποτέ τον κωδικό του.

Σε αντίθεση με άλλη έρευνα (Riley, 2006), ο Riley αναφέρει πως το 52% των χρηστών στη δική του έρευνα δεν έχει αλλάξει ποτέ τον κωδικό του. Το γεγονός ότι η έρευνα του Riley διεξήχθη το 2006, και στην παρούσα έρευνα ένα μικρότερο ποσοστό χρηστών (39%) δεν έχει αλλάξει τον κωδικό του, υποδεικνύει ότι με το πέρασμα του χρόνου οι χρήστες σταδιακά τείνουν να αλλάζουν τον κωδικό τους πρόσβασης συχνότερα.

Στον παρακάτω πίνακα και διάγραμμα παρουσιάζονται τα ποσοστά με την συχνότητα που συνηθίζουν οι χρήστες να αλλάζουν τους κωδικούς τους πρόσβασης.

Πίνακας 5.1: Συχνότητα αλλαγής κωδικού από Έλληνες χρήστες στα ηλεκτρονικά μέσα.

Συχνότητα	Ποσοστό (%)
Ποτέ	39%
Κάθε χρόνο	28%
Κάθε έξι μήνες	14%
Πολύ συχνά	13%
Άλλο	5%

ΑΤΕΙΘ, Τμήμα Πληροφορικής ©2011 Πτυχιακή εργασία Παπαδόπουλος Κυριάκος, Τασιούδης Χρήστος

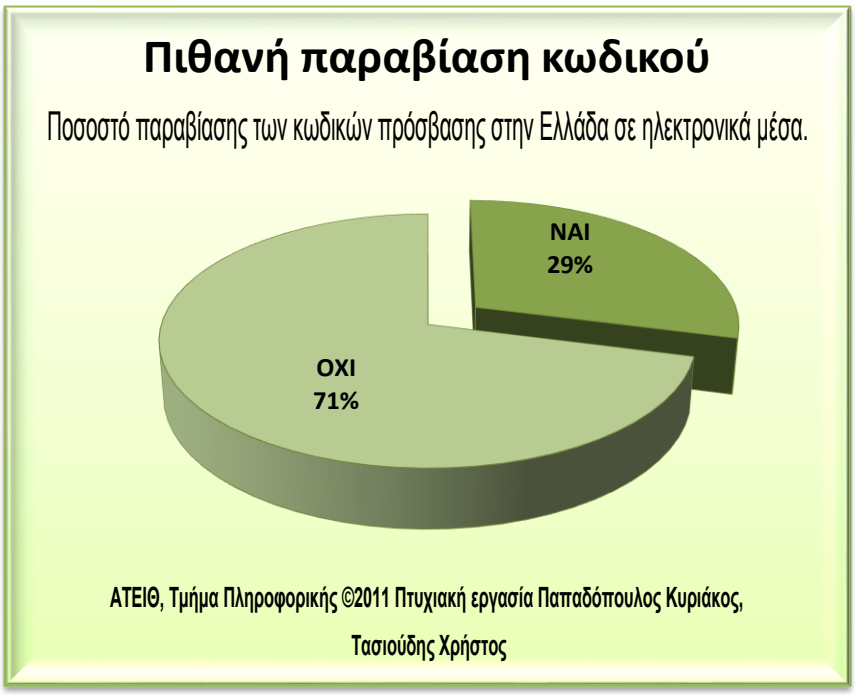
Διάγραμμα 5.2: Συχνότητα αλλαγής κωδικών πρόσβασης στην Ελλάδα σε ηλεκτρονικά μέσα από τους ίδιους τους χρήστες.

Ενδεχομένως οι χρήστες ενημερώνονται σταδιακά για την σημαντικότητα ύπαρξης του κωδικού. Προφανώς κάποιοι είτε έχουν πέσει θύμα κλοπής ή παραβίασης του κωδικού τους και έχουν προχωρήσει σε αλλαγή του, είτε γνωρίζοντας την σπουδαιότητά του για την ασφάλεια των προσωπικών τους δεδομένων, δημιουργούν νέους πιο πολύπλοκους, με αποτέλεσμα να προχωρούν σε συχνές αλλαγές του κωδικού.

Στο σημείο αυτό, αξίζει να αναφερθούν και τα αίτια της αλλαγής ενός κωδικού. Ο κάθε χρήστης μπορεί να αποφασίσει να αλλάξει τον κωδικό του για οποιονδήποτε λόγο. Είτε γιατί θέλει να διατηρεί την ασφάλεια των προσωπικών του στοιχείων αλλάζοντας συχνά τον κωδικό του, είτε γιατί τον έχει αποκαλύψει στον παρελθόν σε τρίτα άτομα. Όπως αναφέρθηκε προηγουμένως όμως, ένας λόγος για τον οποίο ένας χρήστης αποφασίζει να αλλάξει τον κωδικό του είναι η πιθανή παραβίαση του κωδικού του πρόσβασης. Προφανώς κάποιοι χρήστες να έχουν δεχθεί κάποιο e-mail που τους συνιστά την αλλαγή του κωδικού τους ή να έχουν

παρατηρήσει οι χρήστες κάποια συμπεριφορά από την εφαρμογή τους για την οποία δεν έχουν κάνει καμία ενέργεια.

Στην έρευνα μελετάται η πιθανή παραβίαση του κωδικού που μπορεί να υπέστη ένας χρήστης. Έτσι λοιπόν, από την έρευνα εξάγεται το συμπέρασμα πως το 29% των χρηστών πιστεύει πως έχει πέσει θύμα παραβίασης του κωδικού του πρόσβασης, όπως φαίνεται και στο διάγραμμα παρακάτω.



Διάγραμμα 5.3: Ποσοστό παραβίασης των κωδικών πρόσβασης στην Ελλάδα σε ηλεκτρονικά μέσα.

5.5.3. Πολυπλοκότητα που δημιουργούν οι χρήστες.

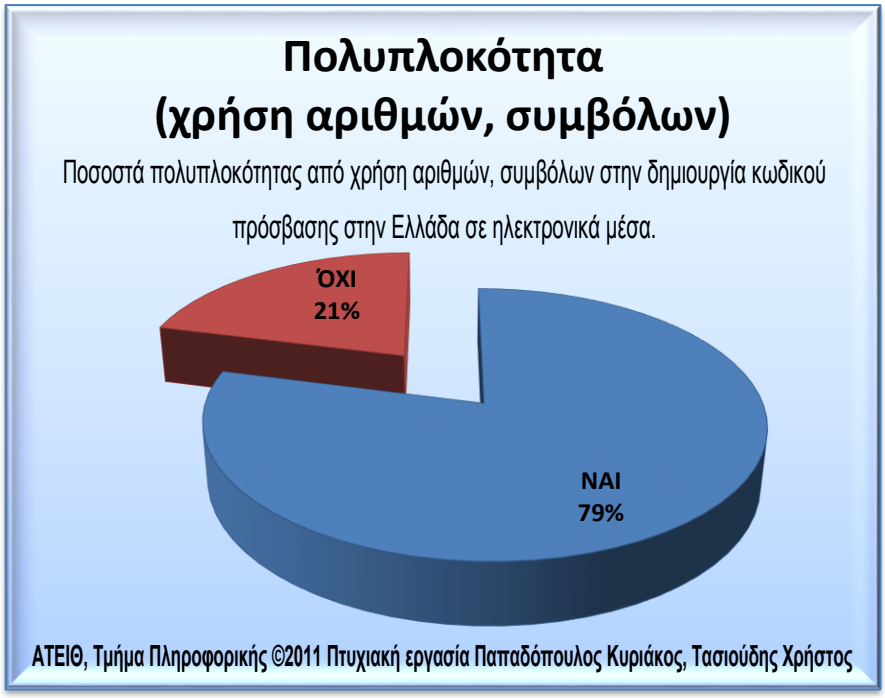
Πέρα από την μυστικότητα του κωδικού, ο μεγαλύτερος παράγοντας που μπορεί να προσδώσει ασφάλεια είναι η πολυπλοκότητα. Η πολυπλοκότητα ενός κωδικού επιτυγχάνεται τόσο από την ποικιλία των χαρακτήρων που απαρτίζουν έναν κωδικό, αλλά εξίσου σημαντικό είναι και το πλήθος των χαρακτήρων που τον αποτελούν. Ένας καλός συνδυασμός του πλήθους χαρακτήρων αλλά και της πολυπλοκότητας αυτών που απαρτίζουν έναν κωδικό, μπορούν να τον μετατρέψουν σε ισχυρό. Η άποψη αυτή τεκμηριώνεται παρακάτω.

Όπως έχει αναφερθεί στο κεφάλαιο 3, είναι προτιμότερο να χρησιμοποιηθεί ως κωδικός πρόσβασης μια απλή έκφραση, όσο εύκολη και αν αυτή μπορεί να θεωρηθεί, παρά ένας μικρός συνδυασμός γραμμάτων και αριθμών (Yan, Blackwell, Anderson, & Grant, 2000). Κωδικοί με πλήθος χαρακτήρων λιγότερο από 8 είναι πολύ εύκολο να παραβιαστούν. Συνεπώς προκύπτει το συμπέρασμα ότι ένας καλός συνδυασμός των παραπάνω παραμέτρων μπορεί να δημιουργήσει έναν ισχυρό κωδικό

Σε παλαιότερη έρευνα (Riley, 2006) δημοσιεύτηκε πως ένα ποσοστό 85,7% χρησιμοποιεί μικρά γράμματα και το 56,5% κάνει χρήση αριθμών έτσι ώστε να αυξήσουν την πολυπλοκότητα. **Στην παρούσα έρευνα, το 79% του δείγματος απάντησε πως κάνει χρήση διαφόρων συμβόλων, γραμμάτων πεζών ή κεφαλαίων, αριθμών ή σημεία στίξης, έναντι 21% που δεν χρησιμοποιεί καθόλου σύμβολα.**

Αν ληφθεί και πάλι υπ' όψιν ότι η έρευνα του Riley διεξήχθη λίγα χρόνια πριν από την παρούσα, εξάγεται το συμπέρασμα ότι το ποσοστό των χρηστών που δεν κάνει χρήση πολυπλοκότητας παραμένει σχεδόν ίδιο. Σε αντίθεση με την ελαφρώς αυξημένη τάση που παρουσιάζουν οι χρήστες να αλλάζουν τον κωδικό τους, δεν πράττουν το ίδιο και για την αύξηση της πολυπλοκότητας. Στην παρούσα έρευνα υπολογίστηκε ότι το 79% των χρηστών κάνουν χρήση αυξημένης πολυπλοκότητας όπως και στην έρευνα του Riley με ποσοστό 85,7%. Αυτοί οι χρήστες προφανώς δημιουργούν πολυπλοκότητα και στους νέους κωδικούς τους. Αντιθέτως, το ποσοστό των χρηστών που δεν έχει κάνει ποτέ χρήση πολυπλοκότητας, ενδεχομένως εξακολουθεί να αποφεύγει μία τέτοια τακτική (χρήσης πολυπλοκότητας), είτε γιατί δεν έχει ενημερωθεί κατάλληλα, είτε γιατί προτιμά να έχει έναν απλό κωδικό όπως έχει αναφερθεί και στο παρελθόν (E. H. Spafford, 1992).

Στο παρακάτω διάγραμμα απεικονίζεται το ποσοστό χρήσης πολυπλοκότητας.



Διάγραμμα 5.4: Ποσοστά πολυπλοκότητας από χρήση αριθμών, συμβόλων στην δημιουργία κωδικού πρόσβασης στην Ελλάδα σε ηλεκτρονικά μέσα.

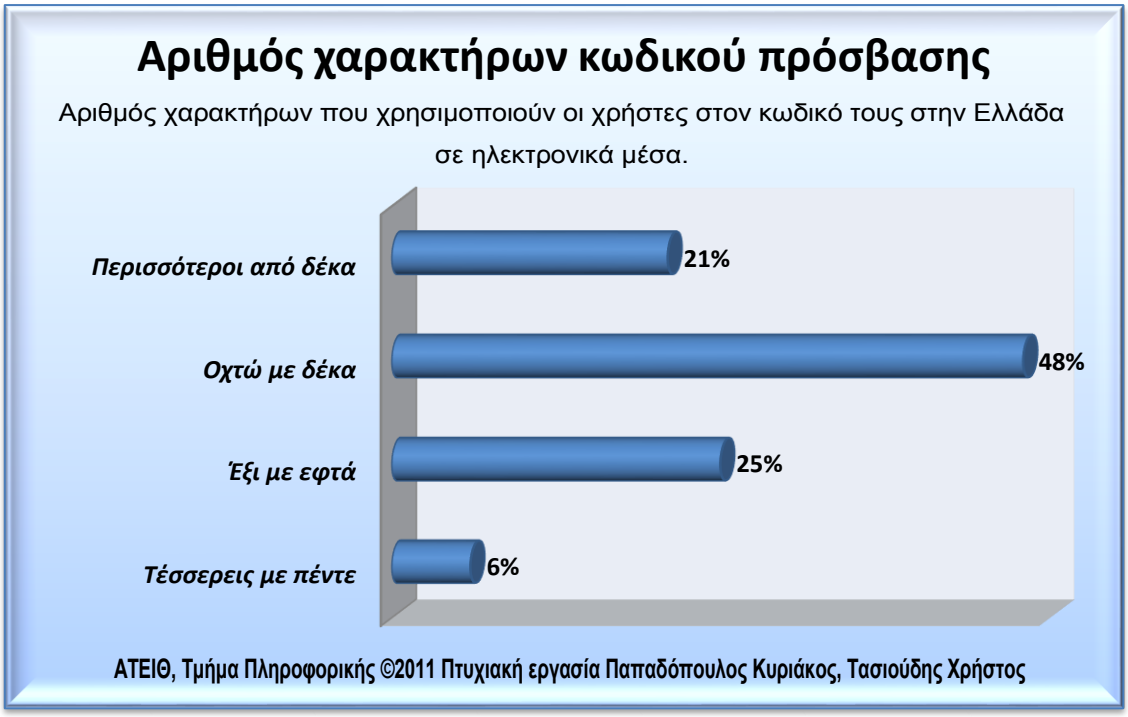
Αναφορικά με το πλήθος των χαρακτήρων που απαρτίζουν τους κωδικούς πρόσβασης του δείγματος, φαίνεται πως το δείγμα κατά πλειονότητα (68%) χρησιμοποιεί περισσότερους από οχτώ χαρακτήρες για τον κωδικό του πρόσβασης.

Στον πίνακα παρακάτω παρουσιάζονται οι επιλογές του δείγματος στον πλήθος των χαρακτήρων που χρησιμοποιούν.

Πίνακας 5.2: Πλήθος χαρακτήρων σε κωδικούς πρόσβασης στην Ελλάδα σε ηλεκτρονικά μέσα.

Πλήθος	Ποσοστό
4-5	6%
6-7	25%
8-10	48%
Περισσότεροι από 10	21%
ΑΤΕΙΘ, Τμήμα Πληροφορικής ©2011 Πτυχιακή εργασία Παπαδόπουλος Κυριάκος, Τασιούδης Χρήστος	

Στο διάγραμμα 5.5 παρακάτω, απεικονίζεται ποσοστιαία ο αριθμός των χαρακτήρων που χρησιμοποιούν οι χρήστες στον κωδικό τους πρόσβασης.



Διάγραμμα 5.5: Αριθμός χαρακτήρων που χρησιμοποιούν οι χρήστες στον κωδικό τους στην Ελλάδα σε ηλεκτρονικά μέσα.

Όπως γίνεται αντιληπτό, ελάχιστο είναι το ποσοστό των ελλήνων χρηστών που χρησιμοποιεί μικρό αριθμό χαρακτήρων για την σύνθεση του κωδικού τους πρόσβασης. Αυτό μπορεί να θεωρηθεί ότι οι συγκεκριμένοι κωδικοί προφανώς δεν έχουν ιδιαίτερη σημασία για τους χρήστες, δηλαδή θεωρούν ότι πιθανή παραβίασή τους δεν θα θέσει σε κίνδυνο προσωπικά τους στοιχεία ή άλλες ευαίσθητες πληροφορίες, κάτι που είναι προφανώς τελείως λανθασμένο.

Αναφορικά με το δείγμα του 6% που κάνει χρήση λιγότερων των 5 χαρακτήρων, ενδεχομένως να φανερώνει ότι οι κωδικοί αυτοί είναι μικροί γιατί απλούστατα ήταν απαραίτητο να δημιουργηθούν, μόνο και μόνο για να έχουν πρόσβαση οι χρήστες σε κάποια εφαρμογή.

Μία άλλη υπόθεση που μπορεί να γίνει είναι ότι 1 στους 3 χρήστες (31%) χρησιμοποιεί λιγότερους από 8 χαρακτήρες. Το συγκεκριμένο ποσοστό χρηστών αγνοεί τους κινδύνους που διατρέχει ο κωδικός του καθώς έχει αναφερθεί ότι κωδικοί λιγότεροι των 8 χαρακτήρων μπορούν πολύ εύκολα να παραβιαστούν με διάφορους μηχανισμούς (Yan et.al., 2000). Επίσης το συγκεκριμένο ποσοστό χρηστών δεν έχει ενημερωθεί κατάλληλα ώστε να μετατρέψει ή να αλλάξει τον

κωδικό του ώστε να είναι μεγαλύτερος των 8 χαρακτήρων και συνεπώς πιο ασφαλής.

Το επίπεδο γνώσης των χρηστών αναφορικά με το πλήθος των χαρακτήρων έχει επίσης ερευνηθεί στην παρούσα έρευνα. Με βάση αυτήν την παράμετρο στην έρευνα, υπάρχει η δυνατότητα να γίνει μια ιδιαίτερα ενδιαφέρουσα σύγκριση μεταξύ της γνώσης για το πλήθος των χαρακτήρων που πρέπει να αποτελούν ένα κωδικό ώστε αυτός να θεωρείται ασφαλής και του αριθμού που ήδη χρησιμοποιούν οι χρήστες. Στο παρακάτω διάγραμμα εμφανίζεται η άποψη του δείγματος σχετικά με το πλήθος των χαρακτήρων που θεωρείται αρκετός για να προσδώσει ασφάλεια.

Διάγραμμα 5.6: Πλήθος χαρακτήρων που θεωρείται αρκετό για την ασφάλεια ενός κωδικού πρόσβασης στην Ελλάδα σε ηλεκτρονικά μέσα.

Όπως γίνεται αντιληπτό από το παραπάνω διάγραμμα, το 66% του δείγματος θεωρεί πως πρέπει να χρησιμοποιεί περισσότερους από 8 χαρακτήρες για να συνθέσει τον κωδικό του πρόσβασης. Συνεπώς,

αντιλαμβανόμαστε ότι 2 στους 3 χρήστες γνωρίζουν ότι ο κωδικός τους δεν είναι ασφαλής αν έχει λιγότερο από 8 χαρακτήρες.

Με μία απλή σύγκριση μεταξύ των διαγραμμάτων 5.5 και 5.6 γίνεται αντιληπτό πως τα ποσοστά είναι σχεδόν ίδια. Το συμπέρασμα που προκύπτει είναι ότι η άποψη που έχει το δείγμα για την ασφάλεια ενός κωδικού μετουσιώνεται και σε πράξη με ελάχιστες αποκλίσεις. Ενδεχομένως ελάχιστα άτομα να μην γνωρίζουν την κρισιμότητα το κωδικού πρόσβασης και να θεωρούν αρκετό και ένα μικρό αριθμό χαρακτήρων. Για την δική τους ασφάλεια όμως χρησιμοποιούν ένα μεγαλύτερο αριθμό χαρακτήρων, γεγονός που αιτιολογείται με την επιλογή για αριθμό χαρακτήρων «τέσσερις με πέντε». Για τη συγκεκριμένη επιλογή, το 6% κάνει χρήση τόσων αριθμών (τέσσερις με πέντε) για τον κωδικό τους, ενώ ένα μεγαλύτερο ποσοστό (8%) θεωρεί πως είναι αρκετό για να την ασφάλεια του κωδικού.

Σε γενικές γραμμές μπορεί να διατυπωθεί η πρόταση ότι, όσο πλήθος χαρακτήρων πιστεύουν οι χρήστες πως απαιτούνται για την ασφάλεια ενός κωδικού, τόσους θα χρησιμοποιήσουν.

5.5.4. Ικανότητα των χρηστών για διάκριση αδύναμων κωδικών.

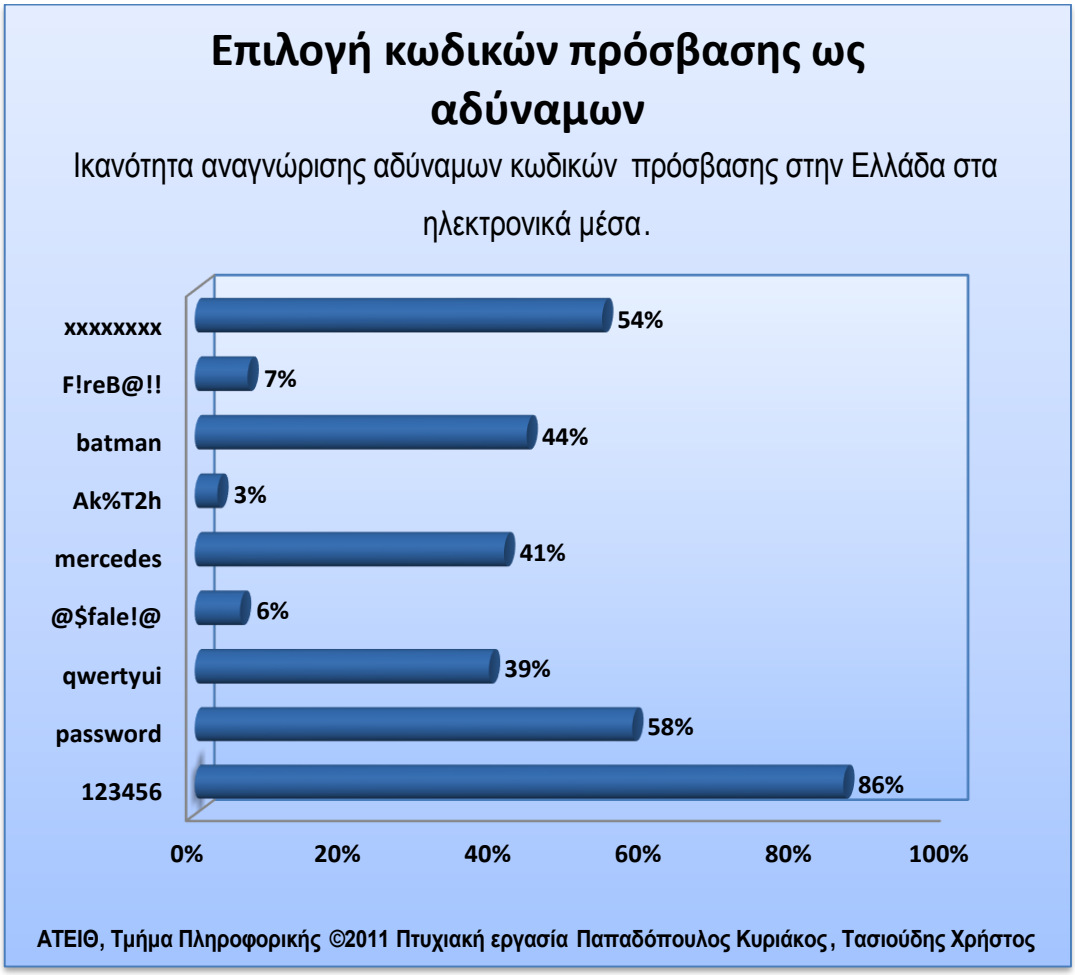
Έχει γίνει αναφορά στο κεφάλαιο 3 για την πολυπλοκότητα των κωδικών πρόσβασης και έχουν παρουσιαστεί ποικίλες τεχνικές για να προστεθεί πολυπλοκότητα σε ένα κωδικό, οι σημαντικότερες εκ των οποίων είναι:

- ✓ το μέγεθός του,
- ✓ η ποικιλία γραμμάτων,
- ✓ χρήση συμβόλων,
- ✓ αντικατάσταση γραμμάτων με κάποιο σύμβολο.

Μέσα από το ερωτηματολόγιο, παρουσιάστηκαν στο δείγμα τυχαίοι κωδικοί πρόσβασης, κάποιοι πολύπλοκοι και κάποιοι που θεωρούνται αδύναμοι, όπως ακολουθιακοί χαρακτήρες ή συνεχόμενοι χαρακτήρες. Από αυτούς που προτάθηκαν αποκαλύφθηκε το ποσοστό του δείγματος που κρίνεται ικανό να διακρίνει αν ένας κωδικός είναι ασφαλής.

Διαπιστώθηκε λοιπόν, πως η μεγαλύτερη μάζα του δείγματος κρίνεται ικανή να διακρίνει έναν ισχυρό κωδικό από αδύναμους. Υπάρχουν όμως και περιπτώσεις χρηστών που αδυνατούν να αντιληφθούν την αδυναμία ενός

κωδικού όπως είναι το "123456". Αναφέρεται χαρακτηριστικά ότι ο συγκεκριμένος κωδικός επιλέχθηκε ως αδύναμος από το 86%, ενώ αναμενόταν ένα ποσοστό μεγαλύτερο του 95%. Αντίστοιχα ο κωδικός "Ak%T2h" που θεωρείται πολύπλοκος επιλέχθηκε ως αδύναμος από το μόλις 3%. Στο διάγραμμα 5.7 παρουσιάζονται αναλυτικά οι υποτιθέμενοι πολύπλοκοι και αδύναμοι κωδικοί με τα αντίστοιχα ποσοστά χρηστών που τους θεώρησαν αδύναμους.



Διάγραμμα 5.7: Ικανότητα αναγνώρισης αδύναμων κωδικών πρόσβασης στην Ελλάδα στα ηλεκτρονικά μέσα.

Αξίζει να σημειωθεί ότι κάποιοι από τους κωδικούς που αναφέρονται στο παραπάνω διάγραμμα είναι εμφανώς αδύναμοι, ενώ κάποιοι άλλοι θεωρούνται πιο ασφαλείς καθώς δείχνουν πιο πολύπλοκοι. Ουσιαστικά, κανείς από τους παραπάνω κωδικούς δεν είναι απόλυτα ασφαλής. Κωδικοί όπως "password", "mercedes" και "batman" θεωρούνται πολύ συνηθισμένες λέξεις και θεωρούνται ιδιαίτερα αδύναμοι κωδικοί. Επίσης οι κωδικοί "123456", "qwerty" και "xxxxxxxx"

ανήκουν στη κατηγορία των κωδικών που περιέχουν επαναλαμβανόμενους ή συνεχόμενους χαρακτήρες και θεωρούνται εύκολο να πληκτρολογηθούν.
Αντιθέτως οι κωδικοί "F!reB@!!", "@$fale!@" είναι σαφώς πιο πολύπλοκοι. Ωστόσο δεν είναι απόλυτα ασφαλείς καθώς αποτελούν μια απλή λέξη στην οποία έχουν αντικατασταθεί ορισμένα γράμματα από κάποια σύμβολα, τεχνική που θεωρείται πλέον γνωστή για μικρό αριθμό χαρακτήρων. Το σπουδαιότερο όμως είναι ότι αυτοί οι κωδικοί ("F!reB@!!", "@$fale!@"), όπως και ο κωδικός "Ak%T2h", αποτελούνται από λιγότερους από 8 χαρακτήρες και όπως προαναφέρθηκε είναι εύκολο να παραβιαστούν.
Αυτό που γίνεται αντιληπτό είναι ότι ο μεγάλος όγκος των χρηστών μπορεί να διακρίνει τους εμφανώς αδύναμους κωδικούς πρόσβασης.

5.5.5. Πλήθος των κωδικών που χρησιμοποιεί ο χρήστης.

Επίσης έχει αναφερθεί στο κεφάλαιο 3 ότι μέσω άλλων ερευνών (‘PasswordResearch.com Statistic, 2011; Riley, 2006) οι οποίες έχουν δημοσιευτεί στο παρελθόν, κάποιοι χρήστες καταλήγουν στην επιλογή να χρησιμοποιούν περισσότερους από έναν κωδικούς ταυτόχρονα για διαφορετικές εφαρμογές. Συγκεκριμένα, στις παραπάνω έρευνες έχει αναφερθεί πως το 70% και το 84% αντίστοιχα, χρησιμοποιούν διαφορετικούς κωδικούς πρόσβασης για τις διάφορες εφαρμογές.
Στην παρούσα έρευνα διαπιστώθηκε πως το 68% του δείγματος χρησιμοποιεί διαφορετικούς κωδικούς πρόσβασης για διαφορετικές εφαρμογές. Προκύπτει το συμπέρασμα πως σε μεγάλο βαθμό η συγκεκριμένη έρευνα προσεγγίζει την έρευνα της ιστοσελίδας «PasswordResearch» με απόκλιση μόλις 2%. Το γενικότερο όμως συμπέρασμα που εξάγεται είναι πως και από τις τρείς αυτές έρευνες αποδεικνύεται ότι ο μεγαλύτερος αριθμός των χρηστών επιλέγει να μην χρησιμοποιεί μόνο έναν κωδικό για τις εφαρμογές του.
Προφανώς σε κάποιες εφαρμογές οι χρήστες χρησιμοποιούν έναν ισχυρό κωδικό πρόσβασης, ενώ σε κάποιες άλλες κάποιον λιγότερο ισχυρό. Αναμφισβήτητα η επιλογή αυτή δεν γίνεται τυχαία. Ενδεχομένως κάποιοι κωδικοί να είναι λιγότερο ασφαλείς με μικρότερη πολυπλοκότητα, σε περιπτώσεις όπου οι χρήστες απαιτούν αρκετές προσβάσεις σε κάποια εφαρμογή, στην οποία όμως λογικά δεν τίθεται

θέμα ασφαλείας των προσωπικών τους στοιχείων (π.χ είσοδος σε κάποιο forum). Επιπροσθέτως, κάποιοι χρήστες προφανώς χρησιμοποιούν 2 ή τρείς κωδικούς και μάλιστα αρκετά ισχυρούς. Ωστόσο δεν επιθυμούν έναν ισχυρό κωδικό να τον χρησιμοποιούν για διαφορετικές εφαρμογές. Πιθανή παραβίαση ή υποκλοπή κάποιου από τους κωδικούς τους πρόσβασης, θα έθετε σε μεγάλο κίνδυνο προσωπικά τους στοιχεία, ενδεχομένως και τραπεζικές συναλλαγές ή οτιδήποτε άλλο ιδιαίτερα σημαντικό για τους ίδιους. Για τον λόγο αυτό για διαφορετικές εφαρμογές χρησιμοποιούν διαφορετικούς κωδικούς και πιθανόν έναν λιγότερο ισχυρό, ίσως τον ίδιο, για ασήμαντες εφαρμογές.

Στο διάγραμμα 5.8 αναπαρίσταται ποσοστιαία το πλήθος των κωδικών που χρησιμοποιεί το δείγμα.

Διάγραμμα 5.8: Πλήθος κωδικών πρόσβασης στην Ελλάδα που χρησιμοποιούν οι χρήστες σε ηλεκτρονικά μέσα.

5.5.6. Τρόποι απομνημόνευσης.

Ένα άλλο στοιχείο το οποίο θεωρείται σημαντικό για την ασφάλεια του κωδικού, όπως έχει αναφερθεί και στο κεφάλαιο 3 (παράγραφος 3.4.2), είναι η μυστικότητα του κωδικού. Ένας κωδικός πρόσβασης από την φύση του θεωρείται μία μυστική

φράση, συνεπώς θα πρέπει να φυλάσσεται έτσι ώστε να μην κινδυνεύει η αποκάλυψή του.

Μέσα από την έρευνα, διαπιστώθηκε πως για την μυστικότητα του κωδικού πρόσβασης το 79% χρησιμοποιεί ως αποθήκευσή του την απομνημόνευση. Αυτό το συμπέρασμα οδηγεί στην υπόθεση ότι οι περισσότεροι χρήστες αισθάνονται ανασφαλείς για την μυστικότητά του κωδικού και προτιμούν να τον γνωρίζουν μόνο οι ίδιοι.

Όπως παρουσιάζεται και στο διάγραμμα 5.9 παρακάτω, οι χρήστες χρησιμοποιούν διάφορες τεχνικές για την αποθήκευση του κωδικού τους. Κατά κύριο λόγο η απομνημόνευση παραμένει ο ασφαλέστερος τρόπος αποθήκευσης, ενώ μέθοδοι όπως η αποθήκευση στον Browser, ή τοπικά στον υπολογιστή του χρήστη ή χειρόγραφα, δεν κερδίζουν την εμπιστοσύνη των χρηστών. Εφόσον αρκετοί χρήστες χρησιμοποιούν περισσότερους από έναν κωδικούς (διάγραμμα 5.8) ενδεχομένως κάποιοι να απομνημονεύονται από αυτούς και κάποιοι λιγότερο σημαντικοί να αποθηκεύονται στον προσωπικό τους υπολογιστή ή χειρόγραφα. Όμως, πιθανή αποθήκευση πέραν της απομνημόνευσης θα αφορά κωδικούς που δεν θα κρύβουν προσωπικά στοιχεία του χρήστη.

Στο διάγραμμα 5.9 παρουσιάζονται οι πιθανοί τρόποι αποθήκευσης των κωδικών πρόσβασης.

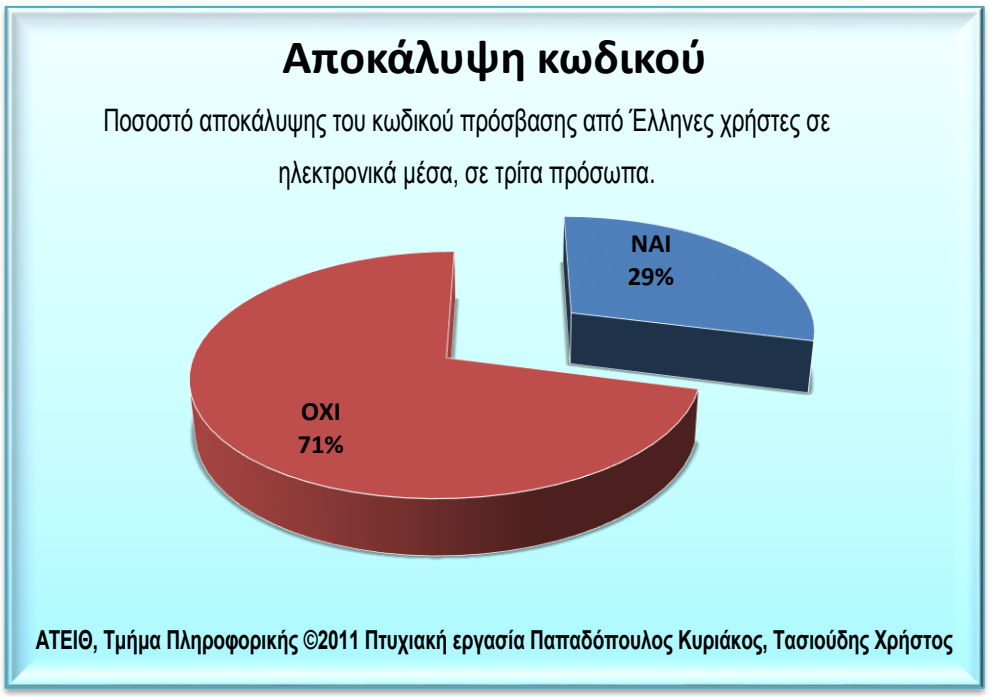
Διάγραμμα 5.9: Τρόποι αποθήκευσης ενός κωδικού πρόσβασης από έλληνες χρήστες σε ηλεκτρονικά μέσα.

5.5.7. Πιθανότητα αποκάλυψης του κωδικού.

Όπως αναφέρθηκε στην προηγούμενη παράγραφο, η μυστικότητα του κωδικού αποτελεί την μεγαλύτερη παράμετρο για την ασφάλειά του. Ωστόσο, για ποικίλους λόγους κάποιοι χρήστες έχουν αποκαλύψει τον κωδικό τους σε τρίτα πρόσωπα για να τους χρησιμοποιήσουν εκ μέρους τους. Φυσικά, όπως έχει γίνει αναφορά και στο κεφάλαιο 4, η χρήση του λογαριασμού κάποιου χρήστη από τρίτα πρόσωπα μπορεί να έχει αρνητικές συνέπειες για τον ίδιο.

Ένας στους δύο χρήστες (46%) απάντησε στην έρευνά ότι έχει αποκαλύψει τον κωδικό του πρόσβασης σε τρίτα άτομα. Το γεγονός αυτό υποδεικνύει ότι οι χρήστες δεν αντιλαμβάνονται την σοβαρότητα της μυστικότητας του κωδικού τους. Θεωρούν προφανώς ότι επειδή γνωρίζουν τα άτομα που αποκαλύπτουν τον

κωδικό τους, δεν θα τους βλάψουν ή ότι δεν θα κάνουν καμία παράνομη ενέργεια εκ μέρους τους.



Διάγραμμα 5.10: Ποσοστό αποκάλυψης του κωδικού πρόσβασης από Έλληνες χρήστες σε ηλεκτρονικά μέσα, σε τρίτα πρόσωπα.

Επίσης στο σύνολο των χρηστών τέθηκε το ερώτημα αν έχουν αποκαλύψει ποτέ τον κωδικό τους στον/ην σύντροφο/σύζυγό τους. **Σε αυτή την περίπτωση φαίνεται πως οι χρήστες είναι πιο επιφυλακτικοί, καθώς μικρότερο ποσοστό έχει αποκαλύψει τον κωδικό του στον/ην σύντροφο/σύζυγό του. Το ποσοστό αυτό αντιστοιχεί στο 34% των χρηστών.**

Στο διάγραμμα 5.11 απεικονίζεται το ποσοστό χρηστών που έχει αποκαλύψει τον κωδικό του πρόσβασης στον/ην σύντροφο/σύζυγό του.

Διάγραμμα 5.11: Ποσοστό αποκάλυψης κωδικού πρόσβασης ηλεκτρονικών μέσων στην Ελλάδα στον/ην σύντροφο/σύζυγό.

Για το συγκεκριμένο θέμα προκύπτει το συμπέρασμα ότι οι χρήστες αισθάνονται περισσότερο τον κίνδυνο να μην αποκαλύψουν στοιχεία στον/ην σύντροφο/σύζυγό τους, από ότι σε άλλα τρίτα πρόσωπα. Ενδεχομένως θεωρούν πως μετά την αποκάλυψη του κωδικού σε τρίτα πρόσωπα, θα μπορέσουν να προχωρήσουν σε αλλαγή του κωδικού τους, ενώ από τον/την σύντροφο/σύζυγό τους δεν μπορούν να αποφύγουν κάτι τέτοιο. Επίσης οι χρήστες δεν θέλουν να παρουσιάσουν προσωπικά τους στοιχεία (π.χ υπόλοιπο λογαριασμού τραπέζης) στον/στην σύντροφο/σύζυγό τους. Ακόμη, πρέπει να σημειωθεί πως μία αποκάλυψη του κωδικού στον/ην σύντροφο/σύζυγο συνεπάγεται συνεχή γνώση του κωδικού από τον/την σύντροφο/σύζυγο, διαφορετικά οποιαδήποτε απόκρυψη του κωδικού στο μέλλον θα θεωρείται ύποπτη ώστε να κρύψει προσωπικά στοιχεία.

Μία ενδεχόμενη αλλαγή του κωδικού και απόκρυψή του από τον/την σύντροφο/σύζυγο πιθανώς να σημαίνει ότι ο χρήστης θέλει να κρύψει πληροφορίες για τον εαυτό του, πέρα από τις περιπτώσεις των τραπεζικών συναλλαγών και σε άλλες εφαρμογές. Τέτοιες περιπτώσεις είναι ο κωδικός του e-mail με το οποίο υπάρχει επικοινωνία τόσο στο πρόγραμμα windows live messenger, αλλά κυρίως στο facebook. Αυτές οι εφαρμογές προσφέρουν την

δυνατότητα στον χρήστη να επικοινωνήσει με άλλα άτομα και γενικότερα να αναπτύξει γνωριμίες. Κάτι τέτοιο σαφώς και θα έθετε σε κίνδυνο την προσωπική σχέση του χρήστη, αν υποθέταμε ότι ο σύντροφος/σύζυγος τους μπορούσε να εισέλθει στο λογαριασμό τους.

Από όλα τα παραπάνω συμπεραίνεται ότι είναι ακόμη λιγότεροι οι χρήστες οι οποίοι αποκαλύπτουν τον κωδικό στον/στην σύντροφο/σύζυγό τους, σε σχέση με τρίτα πρόσωπα, διότι θέλουν να διαφυλάξουν πληροφορίες για την προσωπική τους ζωή.

5.5.8. Κατηγορίες κωδικών που θεωρεί το δείγμα αδύναμες.

Στο διάγραμμα 5.12 εμφανίζονται οι επιμέρους κατηγορίες από τις οποίες κάποιος χρήστης θα μπορούσε να εμπνευστεί για να δημιουργήσει τον κωδικό του πρόσβασης και συνεπώς θεωρούνται αδύναμοι για τους Έλληνες χρήστες.

Διάγραμμα 5.12: Κατηγορίες που εμπνέουν τους Έλληνες χρήστες για δημιουργία κωδικού πρόσβασης σε ηλεκτρονικά μέσα.

Από το παραπάνω διάγραμμα είναι εμφανής η τάση των χρηστών να χρησιμοποιούν ως κωδικό πρόσβασης πράγματα της καθημερινότητας τους και προσωπικά τους στοιχεία. **Έτσι λοιπόν, διαπιστώνεται ότι ένα μεγάλο ποσοστό χρηστών (76%) θεωρεί πως η χρήση του ονόματός τους ως κωδικού θα ήταν κάτι προφανές, ενώ μεγάλο κριτήριο αποφυγής έχουν οι αθλητικές ομάδες που θεωρούνται αδύναμες για ένα χρήστη από το 69% του δείγματος. Επίσης ένας κωδικός μπορεί να θεωρηθεί ευάλωτος αν αυτός είναι το επίθετο ή το όνομα πόλης του χρήστη με ποσοστό 55%.**

5.6. Οι 100 κωδικοί που πρέπει να αποφεύγουν οι Έλληνες χρήστες.

Τέλος, μέσα από την έρευνα, σύμφωνα με τις απαντήσεις του δείγματος (διάγραμμα 5.12) συλλέχθηκαν στοιχεία τα οποία οι Έλληνες χρήστες θεωρούν πολύ κοινά ή εύκολα για την δημιουργία κωδικών πρόσβασης.

Συνεπώς προτείνονται ενδεικτικά 100 κωδικοί πρόσβασης και συνίσταται στους Έλληνες χρήστες να τους αποφεύγουν.

Πρέπει να επισημανθεί ότι τα παρακάτω θα πρέπει να αποφεύγονται τόσο όπως παρουσιάζονται στον πίνακα παρακάτω αλλά και σε συνδυασμό με διάφορους χαρακτήρες που ενδεχομένως να είναι η ημέρα γέννησης του χρήστη, η ομάδα του κ.α. όπως θα αναλυθεί παρακάτω. Η σειρά με την οποία καταγράφονται είναι τυχαία και θεωρείται ότι όλα είναι εξίσου αδύναμα.

Πίνακας 5.3: 100 κωδικοί που πρέπει να αποφύγουν Έλληνες χρήστες.

	TOP 1-25	TOP 26-50	TOP 51-75	TOP 76-100
1	1234	aaaa	QWERTY	Masoutis
2	123456	αααα	qwerty	Marinopoulos
3	4321	ΠΑΟΚ	ΟΣΦΠ	Serres
4	654321	Π.Α.Ο.Κ.	thessaloniki	patra
5	1111	paok-4	athina	Carrefour
6	123	paok4	volos	Multirama
7	321	παοκ-θ4	o-s-f-p	μαρία
8	!@#$	Olympiakos7	A.E.K	Maria
9	$#@!	olympiakos	Pao-13	Giannis
10	password	panathinaikos	osfp	Dimitris
11	!!!!	papadopoulos	giannis	Cyta
12	papanikolaou	giorgos	nikos	panagiwtis
13	katerina	vaso	gewrgia	sofia
14	dimitra	eirini	xristos	xristina
15	bmw	mercedes	mazda	ford
16	fiat	nikolaou	savalas	savidis
17	athanasiadis	dimitriadis	karamanlis	strambopoulos
18	wind	petridis	anna	opel
19	aek-21	eleni	toyota	kostas
20	AEK	natassa	Plaisio	e-shop
21	pao	germanos	vodafone	cosmote
22	Asdf	ασδφ	greece	Conn-x
23	volvo	Q-TELECOM	Tellas	forthnet
24	Audi	honda	nissan	Peugeot
25	Dunlop	michelin	Pirelli	Renault

Έπειτα από την μελέτη του παραπάνω πίνακα των αδύναμων κωδικών πρόσβασης οι παρακάτω οδηγίες-κανόνες θα πρέπει να τηρούνται από τους Έλληνες χρήστες.

Συγκεκριμένα, θα πρέπει να ακολουθηθούν οδηγίες που έχουν αναφερθεί και στην παράγραφο 3.3.5 όπου δίνονται γενικότερες συμβουλές ανεξαρτήτου εθνικότητας του χρήστη. Στους αντίστοιχους ελληνικούς κωδικούς πρόσβασης στα ελληνικά, θα πρέπει επίσης να τηρούνται οι ακόλουθες οδηγίες (Garfinkel & G. Spafford, 1991). Πιο αναλυτικά, πρέπει να αποφεύγεται να χρησιμοποιείται ως κωδικός πρόσβασης ή μέρος αυτού οτιδήποτε από τα ακόλουθα:

- Το όνομά σας,
- Όνομα συζύγου,
- Όνομα πατέρα,
- Όνομα κατοικίδιου,
- Όνομα του παιδιού σας,
- Ονόματα στενών σας φίλων και συνεργατών σας,
- Ονόματα αγαπημένων σας προσωπικοτήτων,
- Όνομα του εργοδότη σας,
- Οποιοδήποτε κύριο όνομα,
- Το όνομα που χρησιμοποιείτε στον υπολογιστή σας,
- Το κινητό σας τηλέφωνο,
- Ο αριθμός των πινακίδων σας,
- Οποιοσδήποτε αριθμός δημοσίου εγγράφου (ταυτότητα, ΑΦΜ, κτλ),
- Οποιαδήποτε ημερομηνία γέννησης συγγενικού προσώπου,
- Άλλες πληροφορίες που μπορούν εύκολα να αντιστοιχηθούν σε εσάς.
- Οποιοδήποτε όνομα χρήστη του υπολογιστή σας σε οποιαδήποτε μορφή,
- Οποιαδήποτε λέξη που υπάρχει στο αγγλικό, ελληνικό και οποιοδήποτε άλλο λεξικό,
- Οποιαδήποτε από τις παραπάνω λέξεις γραμμένη με λατινικούς χαρακτήρες,
- Κωδικός με επαναλαμβανόμενα γράμματα,
- Πρότυπα απλά στο πληκτρολόγιο, όπως qwerty,
- Όλα τα παραπάνω γραμμένα ανάποδα
- Όλα τα παραπάνω προηγούμενα ή ακολουθούμενα από ψηφίο.

Σε αυτό το σημείο παρατίθεται η τάση που έχουν οι χρήστες, σύμφωνα με την έρευνα, να χρησιμοποιούν προσωπικά τους στοιχεία ως κωδικό πρόσβασης. Στο διάγραμμα 5.13 εμφανίζεται η συνήθεια των χρηστών να χρησιμοποιούν προσωπικά στοιχεία ως κωδικό πρόσβασης είτε αυτοτελή είτε με άλλους συνδυασμούς.

Διάγραμμα 5.13: Προσωπικά στοιχεία ως κωδικοί πρόσβασης στη Ελλάδα σε ηλεκτρονικά μέσα.

Στο πίνακα παρακάτω παρουσιάζονται ταξινομημένα τα προσωπικά στοιχεία που χρησιμοποιούνται στους κωδικούς, από το συνηθέστερο προς το πιο σπάνιο.

Πίνακας 5.4: Προσωπικά στοιχεία που χρησιμοποιούνται ως κωδικοί πρόσβασης ελλήνων χρηστών στο διαδίκτυο.

Επιλογές	Ποσοστό
Ημερομηνία γέννησης	31%
Κανένα	28%
Το όνομά μου	26%
Αθλητική ομάδα	14%
Άλλο	10%
Κινητό τηλέφωνο	10%
Διεύθυνση	9%
Μάρκα/μοντέλο αυτοκινήτου	8%
Επωνυμία εταιρείας που εργάζομαι	6%
Όνομα κατοικίδιου ζώου	5%
Όνομα συζύγου	5%
Όνομα παιδιών	5%
Πολιτικό κόμμα	1%
ΑΤΕΙΘ, Τμήμα Πληροφορικής ©2011 Πτυχιακή εργασία Παπαδόπουλος Κυριάκος, Τασιούδης Χρήστος	

Εξάγεται το συμπέρασμα ότι 2 στους 3 χρήστες κάνουν χρήση κάποιου προσωπικού τους στοιχείου, με κάποια πολύ προφανή όπως το όνομά τους και κάποια πιο δύσκολο να προβλεφθούν από τρίτους. Ένα 28% του δείγματος αποφεύγει να χρησιμοποιήσει κάποιο προσωπικό στοιχείο, ενώ το 10% χρησιμοποιεί κάτι διαφορετικό από τα παραπάνω, όπως είναι:

- ✓ ο αριθμός μητρώου επαγγέλματος,
- ✓ το PIN του κινητού τηλεφώνου,
- ✓ αριθμός λογαριασμού τραπέζης,
- ✓ ο τίτλος αγαπημένου τραγουδιού ή ταινίας,
- ✓ κάποια ημερομηνία (π.χ επέτειος),
- ✓ ψευδώνυμο,
- ✓ τροποποιημένο κάποιο από τα e-mail τους,
- ✓ κάποιοι προσωπικοί συνδυασμοί γραμμάτων και λέξεων.

5.7. Προτάσεις δημιουργίας ανθεκτικών κωδικών.

Ως ανθεκτικά- καλά-passwords θεωρούνται αυτά τα οποία (Sanjour, Arensburger, & Brink., 1999):

- Περιέχουν κεφαλαία και μικρά γράμματα.
- Περιέχουν ψηφία και σημεία στίξης.
- Είναι εύκολο να απομνημονευθούν ώστε να μην χρειάζεται να σημειώνονται κάπου.
- Έχουν μήκος πάνω από δέκα χαρακτήρες.
- Να είναι δυνατό να πληκτρολογηθούν γρήγορα ώστε να μην μπορέσει κάποιος να το δει ολόκληρο την ώρα που το πληκτρολογείτε.

Επίσης οι Έλληνες χρήστες θα πρέπει να προσέξουν και τον τρόπο γραφής τους, δηλαδή συνδυασμό μικρών και κεφαλαίων γραμμάτων, αν παρεμβάλλονται από σημεία στίξης αλλά και αν είναι γραμμένα με ελληνικούς ή λατινικούς χαρακτήρες.

- ✓ Είναι δηλαδή πολύ πιθανό κάποιος να έχει ως κωδικό κάποιο προσωπικό του στοιχείο (όνομα, επίθετο, αριθμό κινητού τηλεφώνου, αγαπημένη αθλητική ομάδα κτλ.) στο οποίο να παρεμβάλλεται ανάμεσα στους χαρακτήρες ένα σύμβολο (., -, :, !, @, κτλ). Θα μπορούσε δηλαδή ένας κωδικός να είναι ο "p.a.p.a.d.o.p.o.u.l.o.s"
- ✓ Ακόμη, ενδέχεται κάποιος χρήστης να κάνει στον κωδικό του εναλλαγή κεφαλαίων και πεζών γραμμάτων ή να αντικαθιστά κάποιο γράμμα με σύμβολο, δηλαδή το "α" με "@", το "ε" με "€", το "ξ" με "3", το "θ" με "8", το "ί" με "!" κτλ.

Δεν πρέπει να παραλείπεται το γεγονός ότι οι παραπάνω κωδικοί που θεωρούνται αδύναμοι, δεν πρέπει να χρησιμοποιούνται από Έλληνες χρήστες όχι μόνο όπως παρουσιάστηκαν πρωτύτερα, αλλά και συνοδευόμενοι με κάποιο προσωπικό στοιχείο. Για παράδειγμα κάποιος που έχει ημέρα γέννησης την 16η ημέρα κάποιου μήνα μπορεί να έχει ως κωδικό το όνομα του, να ακολουθεί ένα σύμβολο και έπειτα τον αριθμό της ημέρας γέννησης. Δηλαδή να είναι ο κωδικός "Κωσταντίνος-16" γραμμένος είτε με ελληνικούς είτε λατινικούς χαρακτήρες. Μία απλή επισήμανση είναι ότι ενώ ο κωδικός αυτός είναι ασφαλής, διότι έχει

μεγαλύτερο πλήθος χαρακτήρων από 8, είναι όμως αρκετά προβλέψιμος από κάποιο γνωστό του πρόσωπο.

5.8. Συμπεράσματα

Γενικότερα, από όσα έχουν αναφερθεί παραπάνω, μέσα από την έρευνα διαπιστώθηκε πως υπάρχουν χρήστες οι οποίοι έχουν πλήρη επίγνωση της σοβαρότητας του κωδικού πρόσβασης. Αυτό μάλιστα γίνεται σαφές από τις υποδείξεις αρκετών χρηστών ότι δεν έχουν επιλέξει ποτέ κάποιον από τους αδύναμους κωδικούς που παρατέθηκαν.
Ακόμη αρκετοί χρήστες, αναγνωρίζουν ότι προσωπικά στοιχεία του χρήστη δεν θα πρέπει να αποτελούν τον πλήρη κωδικό τους ή τμήμα αυτού.
Επίσης, πρέπει να τονιστεί ότι το μεγαλύτερο ποσοστό του δείγματος (79%) χρησιμοποιεί την απομνημόνευση ως μέσο αποθήκευσης του κωδικού τους, ενώ αρκετοί (56%) δεν έχουν αποκαλύψει ποτέ τον κωδικό τους σε τρίτα άτομα.
Τέλος, εξήχθη το συμπέρασμα ότι οι χρήστες ανάλογα με τις γνώσεις και ενημέρωση που έχουν λάβει, λειτουργούν αντίστοιχα για τους κωδικούς τους πρόσβασης. Οποιαδήποτε ενημέρωση και μελέτη για τους κωδικούς πρόσβασης από τους χρήστες, χωρίς αμφιβολία θα αποτελέσει ένα θετικό έναυσμα για την ασφάλειά τους.

5.9. ΠΡΟΤΑΣΕΙΣ-ΒΕΛΤΙΩΣΕΙΣ

Μέσα από αυτή την έρευνα βγήκαν αρκετά χρήσιμα συμπεράσματα που αφορούν τους Έλληνες χρήστες. Ωστόσο θα μπορούσε στο μέλλον να διεξαχθεί μία αντίστοιχη έρευνα που να καλύπτει μεγαλύτερο αριθμό χρηστών. Θα πρέπει να βρεθούν τρόποι ώστε ένα αντίστοιχο ερωτηματολόγιο να δημοσιευθεί σε forum, ιστοσελίδες και να προσεγγίσει μεγαλύτερο δείγμα. Ίσως να μπορούσε να συμβάλλει και η Ένωση Πληροφορικών Ελλάδας για την προώθησή του σε εκπαιδευτικές μονάδες. Έτσι θα υπήρχε η δυνατότητα μια τέτοια έρευνα να απευθυνθεί σε εκπαιδευτικούς και μαθητές αλλά κυρίως σε άτομα γνώστες της Πληροφορικής ώστε να παραθέσουν τις δικές τους απόψεις. Ακόμη, σε μία μελλοντική έρευνα θα ήταν χρήσιμες κl ερωτήσεις ανάπτυξης, ζητώντας από το δείγμα να προτείνει συγκεκριμένους αδύναμους κωδικούς για Έλληνες χρήστες.

5.10. ΕΠΙΛΟΓΟΣ

Σε αυτό το κεφάλαιο παρουσιάστηκαν οι διάφορες τεχνικές που χρησιμοποιούν οι Έλληνες χρήστες για την δημιουργία των κωδικών πρόσβασης και οι τρόποι απομνημόνευσης και αποθήκευσης τους. Μελετήθηκε η ικανότητα των χρηστών να διακρίνουν τους ισχυρούς και τους αδύναμους κωδικούς όπου και αποκαλύφθηκε ότι οι περισσότεροι είναι σε θέση να διακρίνουν έναν αδύναμο κωδικό.

Επίσης μέσα από την έρευνα δημιουργήθηκαν και προτάθηκαν οι 100 κωδικοί πρόσβασης που θα πρέπει να αποφεύγουν οι Έλληνες χρήστες.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Γιαννακόπουλος, Χαράλαμπος. 2007. 'Γραμμική - Διαφορική Κρυπτανάλυση και κατασκευή Κρυπτογραφικά ασφαλών S-boxes'. Πανεπιστήμιο Πατρών.

ΕΘΕΓ. (2011). *Εθνικός Θησαυρός Ελληνικής Γλώσσας*. Retrieved May 17, 2011, from http://hnc.ilsp.gr/

Ζορκάδης, Βασίλειος. 2002. *Κρυπτογραφία*. ΠΑΤΡΑ: Ελληνικό Ανοικτό Πανεπιστήμιο.

Κάτος, Β.Α, και Στεφανίδης, Γ.Χ. 2003. *Τεχνικές Κρυπτογραφίας και Κρυπτανάλυσης*. Θεσσαλονίκη: Εκδόσεις ΖΥΓΟΣ.

Λάσκαρη, Έλενα. 2010. 'Κρυπτογραφία και Κρυπτανάλυση με μεθόδους Υπολογιστικής Νοημοσύνης και Υπολογιστικών Μαθηματικών και εφαρμογές'. Πανεπιστήμιο Πατρών.

Μικρός, Γ . (2005). Στατιστικές προσεγγίσεις στην Αυτόματη Κατηγοριοποίηση Κειμένων της ΝΕ. ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ.

Τσιάκης, Θεοδόσης. 2005. 'Η εφαρμοσμένη κρυπτογραφία ως τυπική μέθοδος και μοντέλο για την ασφάλεια των ηλεκτρονικών συναλλαγών'. Retrieved April 4, 2011 (http://phdtheses.ekt.gr/eadd/handle/10442/13784).

Adams, A., Sasse, M. A., & Lunt, P. (1997). Making passwords secure and usable. *People and Computers*, (p.1–20).

Altmann, G. (1988). *Wiederholungen in Texten.* Studienverl Brockmeyer.

Appel, A. W., & Jacobson, G. J. (1988). The world's fastest Scrabble program. *Communications of the ACM*, *31*(5), (p 572–578).

Baillie, W. M.. (1974). Authorship Attribution in Jacobean Dramatic Texts. *Computers in the Humanities*, (p 73–81).

Barton Marthalee, S., & Ben, F. (1984). User-friendly password methods for computer-mediated information systems. *Computers & Security*, *3*(3), (p.186–195).

Beaver, D., J. Feigenbaum, J. Kilian, και P. Rogaway. 1991. 'Security with low communication overhead'. *Advances in Cryptology-CRYPT0'90* (p. 62–76).

Beker, H., & Piper, F. (1982). *Cipher systems: the protection of communications*. Northwood Books.

Bellare, M., και P. Rogaway. 2003. *Introduction to modern cryptography: Lecture notes (2003)*.

Best, K. H. (1998). Results and perspectives of the GoÈttingen project on quantitative linguistics. *Journal of Quantitative Linguistics*, *5*(3), (p 155–162).

Beutelspacher, A. (2005). *Kryptologie*. Vieweg+ Teubner.

Bishop, D. 2003. *Introduction to cryptography with Java applets*. Jones & Bartlett Learning.

BOD, R., & Jennifer, H. A. Y. (2003). *i Stefanie JANNEDY (eds.)(2003): Probabilistic Linguistics*. Cambridge, Massachusetts/London, England: The MIT Press.

Buenaga, M., Gómez, J. M., & Díaz, B. (1997). Using wordnet to complement training information in text categorization. *Proceedings of the Second International Conference on Recent Advances in Natural Language Processing (RANLP)*.

Burnett, S., και S. Paine. 2001. *The RSA Security's Official Guide to Cryptography*. McGraw-Hill, Inc. New York, NY, USA.

Burrows, J. F. (1992). Not Unless You Ask Nicely: The Interpretative Nexus Between Analysis and Information. *Literary and Linguistic Computing*, *7*(2), 91.

Burrows, J. F., & Craig, D. H. (1994). Lyrical drama and the "turbid mountebanks": Styles of dialogue in romantic and renaissance tragedy. *Computers and the Humanities*, *28*(2), 63–86.

Cachin, C. 2005. 'Digital steganography'. *Encyclopedia of Cryptography and Security* (p. 129–168).

Carroll, J. M. (1996). *Computer security*. Butterworth-Heinemann.

Cramer, R., and V. Shoup. 1998. 'A practical public key cryptosystem provably secure against adaptive chosen ciphertext attack'. (p. 13–25) at *Advances in Cryptology—CRYPTO'98*.

De Alvaré, A. M. (1988). *How crackers crack passwords or what passwords to avoid*. Lawrence Livermore National Lab., CA (USA).

Delfs, H., and H. Knebl. 2007. *Introduction to cryptography: principles and applications*. Springer-Verlag New York Inc.

Diffie, W., and M. Hellman. 1976. 'New directions in cryptography'. *Information Theory, IEEE Transactions on* 22(6):644–654.

EAGLES. (1994). *Corpus Encoding DRAFT --- WORK IN PROGRESS*. Retrieved May 17, 2011, from http://xml.coverpages.org/eaglesEncod.html

Federal Information Processing Standards. (2001). . National Institute of Standards and Technology. Retrieved from http://csrc.nist.gov/publications/fips/fips140-2/fips1402.pdf

Forsyth, R. S., & Holmes, D. I. (1996). Feature-finding for test classification. *Literary and Linguistic Computing*, *11*(4), p 163.

Fucks, W. (1956). Die mathematischen Gesetze der Bildung von Sprachelementen aus ihren Bestandteilen. *Nachrichtentechnische Fachberichte*, *3*, (p 7–21).

Garfinkel, S., & Spafford, G. (1991). *Practical UNIX security*.

Gehringer, E. F. (2002). Choosing passwords: security and human factors. *Technology and Society, 2002.(ISTAS'02). 2002 International Symposium on* (p. 369–373).

Goldreich, O. 2001. *Foundations of cryptography: Basic tools*. Cambridge Univ Pr.

Gomez Hidalgo, J. M., & de Buenaga Rodriguez, M. (1997). Integrating a lexical database and a training collection for text categorization. *eprint arXiv: cmp-lg/9709004* (p 9004).

Gordon, A., και A. Jeffrey. 2005. 'Secrecy despite compromise: Types, cryptography, and the pi-calculus'. *CONCUR 2005–Concurrency Theory* (p. 186–201).

Gordon, S. (1995). Social engineering: Techniques and prevention. *Computer Security*.

Goutsos, D. (2010). The Corpus of Greek Texts: a reference corpus for Modern Greek. *Corpora*, *5*(1), (p 29–44).

Grotjahn, R. (1982). Ein statistisches Modell für die Verteilung der Wortl\änge. *Zeitschrift für Sprachwissenschaft*, *1*(1), (p 44–75).

Groza, B., & Petrica, D. (2005). One-time passwords for uncertain number of authentications. *Proceedings of 15 th International Conference on Control Systems and Computer Science, CSCS15*.

Hankerson, D. R, S. A Vanstone, και A. J Menezes. 2004. *Guide to elliptic curve cryptography*. Springer-Verlag New York Inc.

Hatzigeorgiu, N., Gavrilidou, M., Piperidis, S., Carayannis, G., Papakostopoulou, A., Spiliotopoulou, A., Vacalopoulou, A. (2000). Design and Implementation of the online ILSP Greek Corpus. *Proceedings of LREC 2000* (p 1737–1742).

Hatzigeorgiu, N., Mikros, G., & Carayannis, G. (2001). Word length, word frequencies and Zipf's law in the Greek language. *Journal of Quantitative Linguistics*, *8*(3), (p 175–185).

Hitchings, J. (1995). Deficiencies of the traditional approach to information security and the requirements for a new methodology* 1. *Computers & Security*, *14*(5), (p. 377–383).

Hoch, R. (1994). Using IR techniques for text classification in document analysis. *Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval* (p 31–40).

Incarnato, J. S, και W. M Auslander. 2003. *Alphabet soup cryptography*. Google Patents.

Ives, B., Walsh, K. R., & Schneider, H. (2004). The domino effect of password reuse. *Communications of the ACM*, *47*(4), (p. 75–78).

Jakobsen, T., και L. R Knudsen. 1997. 'The interpolation attack on block ciphers'. Σ. 28 στο *Fast Software Encryption: 4th International Workshop, FSE'97, Haifa, Israel, January 1997. Proceedings*.

Jevons, W. S. 1958. 'Principles of Science'. *Daedalus* 87(4):148–154.

Johansson, S., & Hofland, K. (1989). *Frequency analysis of English vocabulary and grammar*. Oxford University Press.

Jorstad, N., και T. S. Landgrave. 1997. 'Cryptographic algorithm metrics'. στο *20th National Information Systems Security Conference*.

Junker, M., & Abecker, A. (1997). Exploiting thesaurus knowledge in rule induction for text classification. *Proceedings of RANLP-97, 2nd International Conference on Recent Advances in Natural Language Processing* (p 202–207).

Kahn, D. 1974. *The codebreakers*. Weidenfeld and Nicolson.

Kapa, S., Hyberger, L., Rea, R. F., & Hayes, D. L. (2007). Complication risk with pulse generator change: implications when reacting to a device advisory or recall. *Pacing and Clinical Electrophysiology*, *30*(6), (p. 730–733).

Karlgren, J. (1999). Stylistic experiments in information retrieval. *Natural language information retrieval*, (p 147–166).

Kaufman, C. W, και S. M Matyas Jr. 1998. *Differential work factor cryptography method and system*. Google Patents.

Kaufman, C., R. Perlman, και M. Speciner. 2002. 'Network security: private communication in a public world'.

Kuo, C., Romanosky, S., & Cranor, L. F. (2006). Human selection of mnemonic phrase-based passwords. *Proceedings of the second symposium on Usable privacy and security* (p. 67–78).

Ledger, G., & Merriam, T. (1994). Shakespeare, Fletcher, and the two noble kinsmen. *Literary and Linguistic Computing*, *9*(3), (p 235).

Leech, G., & others (1996). Recommendations for the Syntactic Annotation of Corpora.

Madnani, N. (2007). Getting started on natural language processing with Python. *Crossroads*, *13*(4), 5–5.

Manning, C. D., Sch\ütze, H., & MITCogNet (1999). *Foundations of statistical natural language processing* (T. 59). MIT Press.

Maurer, U. 2001. 'Cryptography 2000pm10'. (p. 63–85) at *Informatics*.

McEnery, T., Wilson, A., & Barnbrook, G. (1996). Corpus linguistics. *Computational Linguistics*, *24*(2).

Menezes, A. J, P. C Van Oorschot, and S. A Vanstone. 1996. 'Applied Cryptography'. *CRC, Boca Raton*.

Menezes, A. J, P. C Van Oorschot, και S. A Vanstone. 1997. *Handbook of applied cryptography*. CRC.

Merkle, R., and M. Hellman. 1978. 'Hiding information and signatures in trapdoor knapsacks'. *Information Theory, IEEE Transactions on* 24(5):525–530.

Microsoft. (2007). *Microsoft Support*. Retrieved May 20, 2011, from http://support.microsoft.com/kb/189126

Microsoft-what is a password. Retrieved May 20, 2011, from http://windows.microsoft.com/el-GR/windows-vista/What-is-a-password

Mikros G. (2002). *Quantitative linguistics in Greece.* (by Gabriel Altmann, Reinhard Kohler, Raijmund Piotrowski). Berlin: Walter De Gruyter.

Mikros, G., & Carayannis, G. (2000). Modern Greek corpus taxonomy. *Proceedings of the Second International Conference on Language Resources and Evaluation* (p 129–134).

Mikros, G., Hatzigeorgiu, N., & Carayannis, G. (2005). Basic quantitative characteristics of the Modern Greek language using the Hellenic National Corpus. *Journal of Quantitative Linguistics*, *12*(2), (p 167–184).

Miller, G. A., Newman, E. B., & Friedman, E. A. (1958). Length-frequency statistics for written English. *Information and control*, *1*(4), 370–389.

Oxford Dictionaries Online. (2011). *Frequency of the letters of the english alphabet.* Retrieved May 16, 2011, from http://www.oxforddictionaries.com/page/frequencyalphabet

Parker, D. B. (1992). Restating the foundation of information security. *Proceedings of the IFIP TC11, Eigth International Conference on Information Security: IT Security: The Need for International Cooperation* (p. 139–151).

PAROLE. (2011). *Definition of parole by the Free Online Dictionary, Thesaurus and Encyclopedia.* Retrieved May 16, 2011, from http://www.thefreedictionary.com/parole

Password Memorability and Security. (2004). . Cambridge University. Retrieved from http://homepages.cs.ncl.ac.uk/jeff.yan/jyan_ieee_pwd.pdf

PasswordResearch. (2011). Retrieved May 20, 2011, from http://passwordresearch.com/stats/statistic96.html

Pieprzyk, J., T. Hardjono, and J. Seberry. 2003. *Fundamentals of computer security*. Springer Verlag.

Pratt, F. (1942). *Secret and urgent: The story of codes and ciphers*. Blue Ribbon Books.

Rackoff, C., and D. Simon. 1992. 'Non-interactive zero-knowledge proof of knowledge and chosen ciphertext attack'. (p. 433–444) at στο *Advances in Cryptology—CRYPTO'91*.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*(3), (p 191–201).

Riley, S. (2006). Password security: what users know and what they actually do. *Usability News*, *8*(1).

Rivest, R. L, A. Shamir, and L. Adleman. 1978. 'A method for obtaining digital signatures and public-key cryptosystems'. *Communications of the ACM* 21(2):120–126.

Rosenbaum, R., & Fleischmann, M. (2002). Character frequency in multilingual corpus 1–Part 1. *Journal of Quantitative Linguistics*, *9*(3), 233–260.

Rousseau, R., & Zhang, Q. (1992). Zipf's data on the frequency of Chinese words revisited. *Scientometrics*, *24*(2), (p 201–220).

Rudman, J. (1997). The state of authorship attribution studies: Some problems and solutions. *Computers and the Humanities*, *31*(4), 351–365.

Sahami, M., Dumais, S., Heckerman, D., & Horvitz, E. (1998). A Bayesian approach to filtering junk e-mail. *Learning for Text Categorization: Papers from the 1998 workshop* (T. 62).

Sanjour, J., Arensburger, A., & Brink., A. (1999). Choosing a Good Password. Retrieved May 21, 2011, from http://www.cs.umd.edu/faq/Passwords.html

Schneier, B. 1994. 'Description of a new variable-length key, 64-bit block cipher (Blowfish)'. (p. 191–204) at *Fast Software Encryption*.

Schneier, B. 2007. *Applied cryptography: protocols, algorithms, and source code in C*. A1bazaar.

Scott, S., & Matwin, S.. (1999). Feature engineering for text classification. *MACHINE LEARNING-INTERNATIONAL WORKSHOP THEN CONFERENCE-* (p 379–388).

Shannon, C. E. (1951). Prediction and entropy of printed English. *Bell System Technical Journal*, *30*(1), 50–64.

Shannon, C. E. 1949. *Communication theory of secrecy systems*. AT & T.

Singh, S. (1999). Kodboken. *Konsten att skapa sekretess–fraan det gamla Egypten till kvantkryptering. Stockholm: Norstedts Förlag*.

Spafford, E. H. (1992). OPUS: Preventing weak password choices 1. *Computers & Security*, *11*(3), (p. 273–278).

Spearman's Rank Correlation. (2011). Retrieved May 16, 2011, from http://www.angelfire.com/ga2/ibgeography/spearmans.html

Sperberg-McQueen, C. M., & Burnard, L. (1995). The design of the TEI encoding scheme. *Computers and the Humanities*, *29*(1), (p 17–39).

Stallings, W. 2003. 'Cryptography and Network Security: Principles and Practices.' *Practice Hall*.

Stinson, D. R. 2006. *Cryptography: theory and practice*. CRC press.

Summers, W. C., & Bosworth, E. (2004). Password policy: the good, the bad, and the ugly. *Proceedings of the winter international synposium on Information and communication technologies* (p. 1–6).

Tambouratzis, G., Markantonatou, S., Hairetakis, N., & Carayannis, G. (2004). Automatic style categorisation of corpora in the Greek language. *Proceedings of the Second International Conference on Language Resources and Evaluation* (p 135–140).

Těšitelová, M. (1992). *Quantitative linguistics* (T. 37). John Benjamins Pub Co.

van Oorschot, P., and M. Wiener. 2006. 'A known-plaintext attack on two-key triple encryption'. Σ. 318–325 στο *Advances in Cryptology—Eurocrypt'90*.

Veurink, M., M. Koster, and L. T.W.J Berg. 2005. 'The history of DES, lessons to be learned'. *Pharmacy world & science* 27(3):139–143.

Washington, L. C, και W. Trappe. 2002. 'Introduction to cryptography: With coding theory'.

whatsmypass. (2008). *The Top 500 Worst Passwords of All Time*. Retrieved May 20, 2011, from http://www.whatsmypass.com/the-top-500-worst-passwords-of-all-time

Wiedenbeck, S., Waters, J., Birget, J. C., Brodskiy, A., & Memon, N. (2005). PassPoints: Design and longitudinal evaluation of a graphical password system. *International Journal of Human-Computer Studies*, *63*(1-2), (p. 102–127).

Wilcoxon Signed-Rank Test. (2011). Retrieved May 16, 2011, from http://faculty.vassar.edu/lowry/wilcoxon.html

Wimmer, G., & Altmann, G. (1996). The theory of word length: Some results and generalizations. *Glottometrika*, *15*, 112–133.

Wimmer, G., Köhler, R., Grotjahn, R., & Altmann, G. (1994). Towards a theory of word length distribution. *Journal of Quantitative Linguistics*, *1*(1), 98–106.

Yan, J., Blackwell, A., Anderson, R., & Grant, A. (2000). The memorability and security of passwords-some empirical results. *Technical Report-University Of Cambridge Computer Laboratory*.

Zlotkin, G., J. S Rosenschein, Sloan School of Management, και Sloan School of Management. Center for Coordination Science. 1994. *Coalition, cryptography, and stability: Mechanisms for coalition formation in task*

oriented domains. Alfred P. Sloan School of Management, Massachusetts Institute of Technology.

ΠΑΡΑΡΤΗΜΑ Α΄

Σε αυτό το παράρτημα υπάρχουν οι πίνακες με τους χαρακτήρες εκτύπωσης ASCII, καθώς και τους εκτεταμένους χαρακτήρες εκτύπωσης ASCII.

Πίνακας εκτυπώσιμων χαρακτήρων ASCII.

ΔΕΚΑΔΙΚΟΣ	ΧΑΡΑΚΤΗΡΑΣ	ΔΕΚΑΔΙΚΟΣ	ΧΑΡΑΚΤΗΡΑΣ
32	(διάστημα)	80	P
33	!	81	Q
34	"	82	R
35	#	83	S
36	$	84	T
37	%	85	U
38	&	86	V
39	'	87	w
40	(	88	X
41	)	89	Y
42	*	90	Z
43	+	91	[
44	,	92	\
45	-	93	]
46	.	94	^
47	/	95	_
48	0	96	`
49	1	97	a
50	2	98	b
51	3	99	c
52	4	100	d

53	5	101	e
54	6	102	f
55	7	103	g
56	8	104	h
57	9	105	i
58	:	106	j
59	;	107	k
60	<	108	l
61	=	109	m
62	>	110	n
63	?	111	o
64	@	112	p
65	A	113	q
66	B	114	r
67	C	115	s
68	D	116	t
69	E	117	u
70	F	118	v
71	G	119	w
72	H	120	x
73	I	121	y
74	J	122	z
75	K	123	{
76	L	124	|
77	M	125	}
78	N	126	~
79	O	127	DEL

Πίνακας εκτεταμένων χαρακτήρων εκτύπωσης ASCII.

ΔΕΚΑΔΙΚΟΣ	ΧΑΡΑΚΤΗΡΑΣ	ΔΕΚΑΔΙΚΟΣ	ΧΑΡΑΚΤΗΡΑΣ
128	Ç	192	└
129	ü	193	┴
130	é	194	┬
131	â	195	├
132	ä	196	─
133	à	197	┼
134	å	198	╞
135	ç	199	╟
136	ê	200	╚
137	ë	201	╔
138	è	202	╩
139	ï	203	╦
140	î	204	╠
141	ì	205	═
142	Ä	206	╬
143	Å	207	╧
144	É	208	╨
145	æ	209	╤
146	Æ	210	╥
147	ô	211	╙
148	ö	212	Ô
149	ò	213	╒
150	û	214	╓
151	ù	215	╫

152	ÿ	216	╪
153	Ö	217	┘
154	Ü	218	┌
155	¢	219	█
156	£	220	▄
157	¥	221	▌
158	Pts	222	▐
159	*ƒ*	223	▀
160	á	224	α
161	í	225	ß
162	ó	226	Γ
163	ú	227	π
164	ñ	228	Σ
165	Ñ	229	σ
166	ª	230	µ
167	º	231	τ
168	¿	232	Φ
169	⌐	233	Θ
170	¬	234	Ω
171	½	235	δ
172	¼	236	∞
173	¡	237	φ
174	«	238	ε
175	»	239	∩
176	░	240	≡
177	▒	241	±
178	▓	242	≥

179	┤	243	≤
180	╡	244	⌠
181	╢	245	⌡
182	╖	246	÷
183	╕	247	≈
184	╣	248	≈
185	║	249	∙
186	╗	250	∙
187	╝	251	√
188	╜	252	ⁿ
189	╛	253	²
190	┐	254	■
191	└	255	

ΠΑΡΑΡΤΗΜΑ Β΄

Σε αυτό το παράρτημα περιλαμβάνεται το λογισμικό ανάπτυξης εφαρμογής του κεφαλαίου 3. Συγκεκριμένα υπάρχει κώδικας σε γλώσσα **Java** που αναπτύχθηκαν στο προγραμματικό περιβάλλον του **NetBeans. Η εφαρμογή αυτή έχει ως στόχο να ανιχνεύσει την αρχική γλώσσα προέλευσης του κρυπτοκειμένου, δεδομένου όμως ότι αυτό θα προέρχεται από την αγγλική ή την ελληνική γλώσσα.**

```
package caesarcypher;

//EISAGWGH PAKETWN

import java.io.*;
import java.io.BufferedReader;
import java.io.File;
import java.io.FileReader;
import java.io.FileNotFoundException;
import java.io.IOException;
import java.util.logging.Level;
import java.util.logging.Logger;
import java.util.Random;

public class CaesarCypher {


  //DILWSH METABLHTWN

  static Writer output = null;          /* dilwnoume ws arxiki timi
                                        tou output iso me "null"
                                         */
  static File file = new File("Ceasar.txt");   /* Dimiourgeite to
                                                  Caesar.txt sti metavliti
                                                 file apo to opoio tha diavazei
```

```
                                                    ena keimeno
                                                          */
static File file2 = new File("Encrypted.doc"); /* Dimiourgeite
                                          to arxeio   Encrypted.txt sti metavliti
                    file2 sto opoio kai tha apothikeuetai to kruptografimeno keimeno
                                                          */

static StringBuffer contents = new StringBuffer();
static BufferedReader reader = null;        /* Dilwnoume to reader
                                                    iso me "null" */
static String dokimi = null;              /* Dilwnoume mia metavliti
                                        dokimi tupou String isi me "null" */
static String test="";              /* Dilwnoume mia metavliti test
                                                    tupou String */
static Random generator = new Random();    /* Dilwnoume tin
                                              random timi  generator */


public static void main(String[] args) throws IOException {


      try {

          reader=new  BufferedReader(new  FileReader(file));
                                    /* To "reader" diavazei apo to file
                                          sugkekrimena to Caesar.txt
                                    */

    String text = null;
    /*Efoson den einai adeio to Caesar.txt*/
    while ((text = reader.readLine()) != null) {
```

```
contents.append(text).append(System.getProperty("line.separat
or"));
```

//pairnei to String tis platformas gia na markarei to telos kathe grammis gia na to xrisimopoihsei argotera

```
        }

        } catch (FileNotFoundException e) { //se periptwsi
pou leipei to arxeio na petaksei eksairesi

            e.printStackTrace();

          } catch (IOException e) {

            e.printStackTrace();

            } finally {

          try {

                if (reader != null) {

                reader.close();

                }

                } catch (IOException e) {

                e.printStackTrace();

                }

                }
```

```
                    dokimi=contents.toString();     /*h      metavliti
        dokimi pairnei ta periexomena tou Caesar.txt ws sumvoloseires
                                                                */

    int r = generator.nextInt(33); /*stin akeraia timi "r" orizoume
                        mia tuxaia timi eurous 33
                       */
    int r2 = generator.nextInt(33); /*stin akeraia timi "r2" orizoume
                        mia tuxaia timi eurous 33
                       */
    final int MOVE_UP = r;  /* Orizoume stin metavliti "MOVE_UP" tin
                      metatopisi r*/
     final int MOVE_UP2 = r2; /*Orizoume stin metavliti "MOVE_UP2" tin
                          metatopisi r2*/
     String plainText = dokimi;
     char character = 0;

    System.out.print("Encrypted sentence is: " );

        /*Ksekiname apo 1 kai ftanoume ews -2 etsi wste na apofugoume tin
        ektupwsi kapoiwn peritwn sumvolwn sto kruptofrafimeno keimeno stin
            arxi k sto telos tou */

    for(int    iteration=1;    iteration<plainText.length()-2;
iteration++)
       {
                character = plainText.charAt(iteration);
                    /*pairnei        xaraktires tou plaintext*/

   if ((int)character<122){ /* Me vasi ton pinaka ASCCII oi
            agglikoi xaraktires vriskontai katw apo tin thesi 122*/

   if ((int)character>84 && (int)character<106){
```

```
            character = (char) (character + (MOVE_UP + 130));
```

/*Ginete metatopisi kata 130 theseis giati ta kefalaia grammata einai pio panw apo mikra, opote de theloume na ginei metatopisi enos kefalaiou grammatos k na antikatastathei apo ena mikro tis idias glwssas

*/

```
                        }

                        else if ((int)character>106 &&
(int)character<123){

                            character = (char) (character +
(MOVE_UP + 155));
                            /*Ginete metatopisi kata 155
theseis*/
                        }


     else {
          character = (char) (character + (MOVE_UP+80));
```

/*Ginetai i metatopisi kata 80 theseis*/

```
                    }

                    }

     else{
          character = (char) (character - (MOVE_UP2+100));
```

/*Ginetai i metatopisi kata 100 theseis*/

```
           }
                    test=test+character;
```

/* To kruptografimeno pou prokuptei prostithetai kathe fora sto telos tou to neo kruptografimeno sumvolo

*/

```
        output = new BufferedWriter(new FileWriter(file2));

        output.write(test);

          System.out.print(character);
                        }
           output.close();   /* Kleinei to arxeio eksodou*/

          System.out.println();

         if ((int)character>322) /*Sugkrinoume ton xaraktira an einai
                                          megaluteros apo tin timi 322
                                  */
            System.out.println("To keimeno einai elliniko");
        /* An   einai alithis o elenxos tote to keimeno mas tha einai elliniko*/
           else
              System.out.println("To keimeno einai aggliko");
        /* An einai pseudis o elenxos tote to keimeno mas tha einai aggliko*/
      }
}
```

ΠΑΡΑΡΤΗΜΑ Γ΄

Σε αυτό το παράρτημα περιλαμβάνεται η δεύτερη εφαρμογή λογισμικού που αναπτύχθηκε με σκοπό να πραγματοποιήσει την διαδικασία της κρυπτογράφησης, με επιπλέον χαρακτηριστικό να μελετήσει την επιρροή που έχει ένας αλγόριθμος κρυπτογράφησης στην συχνότητα και κατανομή των γραμμάτων.

```
//dilwsi paketwn

import java.io.BufferedReader;
import java.io.BufferedWriter;
import java.io.InputStream;
import java.io.OutputStream;
import java.io.FileInputStream;
import java.io.FileOutputStream;
import java.io.File;
import java.io.FileWriter;
import javax.crypto.Cipher;
import javax.crypto.SecretKey;
import javax.crypto.spec.IvParameterSpec;
import javax.crypto.CipherInputStream;
import javax.crypto.CipherOutputStream;
import javax.crypto.KeyGenerator;
import java.io.IOException;
import java.io.InputStreamReader;
import java.io.Reader;
import java.io.Writer;
import java.nio.charset.Charset;
import java.security.spec.AlgorithmParameterSpec;
import java.util.Scanner;
import java.util.logging.Level;
import java.util.logging.Logger;

public class JavaApplication1
```

```
{
     /*DILWNONTAI OI KRUPTALGORITHMOI ecipher dcipher gia kruptografisi
                    kai apokruptografisi*/
  Cipher ecipher;
   Cipher dcipher;
   static Writer output = null;  /*dilwnetai arxika i eksodos null sto
                                   output*/
   static File file = new File("pososta.doc");  /*ftiaxnoume to
            arxeio eksodou pou krataei ta pososta grammatwn */
   static StringBuilder builder = new StringBuilder();

   public JavaApplication1(SecretKey key)
        {
        // Create an 8-byte initialization vector
         byte[] iv = new byte[]
             {
              0x00, 0x01, 0x02, 0x03, 0x04, 0x05, 0x06, 0x07,
0x08, 0x09,0x0a, 0x0b, 0x0c, 0x0d, 0x0e, 0x0f
              };
//ews kai tin methodo apo-kryptografisis tou aes, to akoloutho tmima kwdika
apotelei autoteles kommati tou algorithmou aes
    AlgorithmParameterSpec paramSpec = new
IvParameterSpec(iv);
                try
                {
     ecipher= Cipher.getInstance("AES/CBC/PKCS5Padding");

     dcipher=Cipher.getInstance("AES/CBC/PKCS5Padding");

     ecipher.init(Cipher.ENCRYPT_MODE, key, paramSpec);
     dcipher.init(Cipher.DECRYPT_MODE, key, paramSpec);
                }
                catch (Exception e)
                {
```

```
                    e.printStackTrace();
                }
        }
```

// Buffer used to transport the bytes from one stream to another
byte[] buf = new byte[1024];

//methodos kruptografisis tou aes

```
public void encrypt(InputStream in, OutputStream out)
    {
            try
            {
```

// Bytes written to out will be encrypted

```
        out = new CipherOutputStream(out, ecipher);
```

// Read in the cleartext bytes and write to out to encrypt

```
        int numRead = 0;
                    while ((numRead = in.read(buf)) >= 0)
                    {
                            out.write(buf, 0, numRead);
                    }
                    out.close();
            }
            catch (java.io.IOException e)
            {
            }
    }
```

//**methodos apokruptografisis tou aes**

```
public void decrypt(InputStream in, OutputStream out)
   {
       try
          {
```

// Bytes read from in will be decrypted

```
            in = new CipherInputStream(in, dcipher);

        // Read in the decrypted bytes and write the cleartext to out
                        int numRead = 0;
                        while ((numRead = in.read(buf)) >= 0)
                        {
                                out.write(buf, 0, numRead);
                        }
                        out.close();
                }
                catch (java.io.IOException e)
                {
                }
        }

        public static String readFile(String DESTest, String
csName)
            throws IOException {
    Charset cs = Charset.forName(csName);
    return readFile(DESTest, cs);
}

    public static String readFile(String DESTest, Charset cs)
            throws IOException {
  // No real need to close the BufferedReader/InputStreamReader
  // as they're only wrapping the stream
  FileInputStream stream = new FileInputStream(DESTest);
    try {
        Reader     reader     =     new     BufferedReader(new
InputStreamReader(stream, cs));

        char[] buffer = new char[8192];
        int read;
```

```
        while ((read = reader.read(buffer, 0, buffer.length))
> 0) {
            builder.append(buffer, 0, read);
        }
         String selectString = builder.toString();
     return builder.toString();

    } finally {
    // Potential issue here: if this throws an IOException,
    // it will mask any others. Normally I'd use a utility
    // method which would log exceptions and swallow them
        stream.close();
    }

}


    public static void main(String args[])
        {
             try
             {
            // Generate a temporary key. In practice, you would save this key.
            // See also e464 Encrypting with DES Using a Pass Phrase.

  KeyGenerator kgen = KeyGenerator.getInstance("AES");
  kgen.init(128);
  SecretKey key  = kgen.generateKey();

            // Create encrypter/decrypter class
   JavaApplication1 encrypter = new JavaApplication1(key);

            // Encrypt
   encrypter.encrypt(new   FileInputStream("DESTest.txt"),new
FileOutputStream("Encrypted.txt"));
```

```
            // Decrypt
  encrypter.decrypt(new  FileInputStream("Encrypted.txt"),new
FileOutputStream("Decrypted.txt"));
                }
        catch (Exception e)
                {
                        e.printStackTrace();
                }
```

//dilwnontai oi pinakes grammatwn gia kefalaia kai mikra ellinika kai agglika

```
       char[] capital = { 'A', 'B', 'C', 'D', 'E', 'F', 'G',
'H', 'I', 'J','K', 'L', 'M', 'N',
'O', 'P', 'Q', 'R', 'S', 'T', 'U', 'V', 'W', 'X', 'Y', 'Z',
'A', 'B', 'Γ', 'Δ', 'E', 'Z', 'H', 'Θ', 'I', 'K', 'Λ', 'M',
                'N', 'Ξ', 'O', 'Π','P', 'Σ', 'T', 'Y', 'Φ',
'X', 'Ψ', 'Ω'};

        char[] small = { 'a', 'b', 'c', 'd', 'e', 'f', 'g',
'h', 'i', 'j', 'k', 'l', 'm', 'n',
'o', 'p', 'q', 'r', 's', 't', 'u', 'v', 'w', 'x', 'y', 'z',
'α', 'β', 'γ', 'δ', 'ε', 'ζ', 'η', 'θ', 'ι', 'κ', 'λ', 'μ',
                'ν', 'ξ', 'ο', 'π','ρ', 'ς', 'τ', 'υ', 'φ',
'χ', 'ψ', 'ω'};

        Scanner scan;  //dilwnetai enas anixneutis me onoma scan


        try {
           scan          =          new          Scanner(new
File("C:\\Users\\CHRIS\\Desktop\\Ptyxiaki\\DESTest.txt"));
/* o scan psaxnei sto arxeio DESTest.txt,
```

```
  prepei na dwthei i swsti diadromi pou vriksetai  to arxeio
*/
        } catch (Exception e) {

                 System.out.println("File not found");
                return;
                 }
                try {
                  output     =     new     BufferedWriter(new
FileWriter(file));
              // stin eksodo grafetai to arxeio file, diladi ta pososta.txt

                 int[] count = new int[50];
                // dilwnoume ton pinaka 50 thesewn 26 agglika + 24 ellinika



                 while(scan.hasNextLine()) {
                // oso uparxei epomenos xaraktiras, pou vriksei o scan

                 String line = scan.nextLine();

                 System.out.println("Line read: " + line);
                 output.write("Line read: " + line+'\n');

                 char[] digit = line.toCharArray();
                /* dilwnetai o pinakas xaraktirwn digit o opoios
                metatrepei ta gramamta eisodou se xaraktires */

                  output.write('\n');
                   System.out.print('\n');

          for(int i = 0; i < digit.length; i++) {
              // eksetazoume oso to plithjos twn gramamtwn eksantlithei
         for(int j = 0; j < 50; j++) {
```

```
//kai oso uparxoun xaraktires apo autous pou exoun dilwthei
if(digit[i] == capital[j] || digit[i] == small[j]){
//stous pinakes "capital" kai "small"
count[j]++;

        break;
        }
        }
        }
        }

        for (int i = 0; i < 50; i++)
        {

        System.out.print(" " + capital[i]);
        output.write(" " + capital[i]);
        System.out.println(" " + count[i]);
        output.write(" " + count[i]+'\n');

        }



        output.close();

        }


        catch (IOException ex) {

Logger.getLogger(JavaApplication1.class.getName()).log(Level.
SEVERE, null, ex);

        }
```

```
        }
}
```

Η εφαρμογή αυτή, όπως αναφέρθηκε και στο κεφάλαιο 3, αποτελείται από 2 ξεχωριστές εφαρμογές έτσι ώστε να μην υπάρχει πρόβλημα με την αναγνώριση των γραμμάτων κατά την διαδικασία εισαγωγής του κρυπτογραφημένου κειμένου "Encrypted.txt". Έτσι λοιπόν, εφ' όσον έχει μετατραπεί το κρυπτογραφημένο κείμενο σε utf-8 όπως έχει υποδειχθεί στο κεφάλαιο 3, ακολουθεί το παρακάτω κομμάτι κώδικα το οποίο ολοκληρώνει την διαδικασία της αποκρυπτογράφησης.

```
package javaapplication2;
//dilwsi pakerwn
import java.io.BufferedReader;
import java.io.BufferedWriter;
import java.io.FileNotFoundException;
import java.io.InputStream;
import java.io.OutputStream;
import java.io.FileInputStream;
import java.io.FileOutputStream;
import java.io.File;
import java.io.FileWriter;
import javax.crypto.Cipher;
import javax.crypto.SecretKey;
import javax.crypto.spec.IvParameterSpec;
import javax.crypto.CipherInputStream;
import javax.crypto.CipherOutputStream;
import javax.crypto.KeyGenerator;
import java.io.IOException;
import java.io.InputStreamReader;
import java.io.Reader;
import java.io.Writer;
import java.nio.charset.Charset;
```

```
import java.security.spec.AlgorithmParameterSpec;
import java.util.Scanner;
import java.util.logging.Level;
import java.util.logging.Logger;



public class JavaApplication2

        static Writer output = null;
        static File file = new File("pososta.doc"); /* dilwnetai to arxeio pososta, sto opoio tha emfanizontai taksinomimena ta grammata me ton aukson airhmo emfanisis sto arxeio eisodou*/

        public static void main(String[] args) throws FileNotFoundException {

            /*dilwnontai episis oi pinakes capital k small gia ta kefalaia kai mikra grammata tis ellinikis kai aggliki alfabitou*/
        char[] capital = { 'A', 'B', 'C', 'D', 'E', 'F', 'G', 'H', 'I', 'J','K', 'L', 'M', 'N',
'O', 'P', 'Q', 'R', 'S', 'T', 'U', 'V', 'W', 'X', 'Y', 'Z',
'A', 'B', 'Γ', 'Δ', 'E', 'Z', 'H', 'Θ', 'I', 'K', 'Λ', 'M',
                'N', 'Ξ', 'O', 'Π','P', 'Σ', 'T', 'Y', 'Φ', 'X', 'Ψ', 'Ω'};

         char[] small = { 'a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j', 'k', 'l', 'm', 'n',
'o', 'p', 'q', 'r', 's', 't', 'u', 'v', 'w', 'x', 'y', 'z',
'α', 'β', 'γ', 'δ', 'ε', 'ζ', 'η', 'θ', 'ι', 'κ', 'λ', 'μ',
                'ν', 'ξ', 'ο', 'π','ρ', 'ς', 'τ', 'υ', 'φ', 'χ', 'ψ', 'ω'};

        Scanner scan;  //dilwnetai enas anixneutis me onoma scan
```

```
        try {
```

/*dilwnetai i diadromi gia tin euresi tou arxeiou Encrypted.txt wste na mporesei auto na eisaxthei */

```
                scan = new Scanner(new File("C:\\Users\\CHRIS\\Desktop\\Ptyxiaki\\Encrypted.txt"));



                } catch (Exception e) {

                System.out.println("File not found");
```

/*an den vrethei to arxei ototе ektupwnetai to antistoixo minuma*/

```
                return;

                }

                try {

                output = new BufferedWriter(new FileWriter(file));

                int[] count = new int[50];
```

/*dilwnetai enaw pinakas me onoma count 50 thesewn*/

```
                while(scan.hasNextLine()) {
```

/*ginetai elegxos, oso o anixneutis "scan" entopizei epomeni grammi na kanei tous akolouthous elegxous*/

```
                String line = scan.nextLine();
```

/*stin metavliti "line" kratatai i epomeni grammi pou entopizei o anixneutis*/

```
                System.out.println("Line read: " + line);
/*ektupwnetai i kathe grammi pou diavazei o scan apo to arxeio
Encrypted.txt*/
                output.write("Line read: " + line+'\n');

                char[] digit = line.toCharArray();

                output.write('\n');
                System.out.print('\n');

                for(int i = 0; i < digit.length; i++) {
                for(int j = 0; j < 50; j++) {
                if(digit[i] == capital[j] || digit[i] ==
small[j]) {
                count[j]++;

                break;
                }
                }
                }
                }

                for (int i = 0; i < 50; i++)
                {

                System.out.print(" " + capital[i]);
                output.write(" " + capital[i]);
                System.out.println(" " + count[i]);
                output.write(" " + count[i]+'\n');

                }
```

```
                output.close();//kleinei to arxeio eksodou

                }


                catch (IOException ex) {

Logger.getLogger(JavaApplication2.class.getName()).log(Level.
SEVERE, null, ex);

                }

        }
}
```

ΠΑΡΑΡΤΗΜΑ Δ΄

Σε αυτό το παράρτημα παρουσιάζεται το περιεχόμενο του ερωτηματολογίου με τις απαντήσεις κάθε ερώτησης από το δείγμα των 594 χρηστών. Το ερωτηματολόγιο βρίσκεται στον ακόλουθο σύνδεσμο:
http://tinyurl.com/pass-questions
Περιλαμβάνονται οι ερωτήσεις με στατιστικά δεδομένα αλλά και ταξινομημένα με χρήση πινάκων. Επισημαίνεται ότι ακόμη ερωτηματολόγιο παραμένει ακόμη ενεργό και στο μέλλον θα μπορούσε να επεκταθεί, είτε με επιπρόσθετες ερωτήσεις, είτε με απαντήσεις περισσότερων χρηστών.
Για το ερωτηματολόγιο χρησιμοποιήθηκε το Google.docs, στο οποίο υπάρχει η δυνατότητα ενημέρωσης των δεδομένων έπειτα από λήψη περισσότερων απαντήσεων και τροποποίηση των δεδομένων. Η εξαγωγή στατιστικών αποτελεσμάτων γίνεται αυτόματα από την εφαρμογή. Η δημιουργία του και ενημέρωσή του έγινε σε συνεργασία με τον επιβλέποντα καθηγητή της έρευνας αλλά και από το δείγμα που έλαβε μέρος σε αυτήν.

Ερώτηση 1:

Χρησιμοποιείτε ή έχετε χρησιμοποιήσει ποτέ κάποιο από τα παρακάτω password (ή κάποιο που να μοιάζει πάρα πολύ με αυτά);

Πίνακας 5.5: Δεδομένα πρώτης ερώτησης.

Ποσοστά χρήσης από κάποιον από τους παρακάτω κωδικούς στο παρελθόν		
Επιλογές	**Απαντήσεις**	**Ποσοστό**
Όχι, δεν χρησιμοποιώ	150	32%
1234	124	27%
Το όνομά μου	124	27%
123	64	14%
παοκ (ελληνικά ή λατινικά)	62	13%
1111	39	8%
Το επώνυμό μου (ελληνικά ή λατινικά)	39	8%

Επιλογές	Απαντήσεις	Ποσοστό
4321	35	8%
asdf	30	6%
Άλλη αθλητική ομάδα	27	6%
password	20	4%
!@#$	18	4%
οσφπ (ελληνικά ή λατινικά)	16	3%
αεκ (ελληνικά ή λατινικά)	15	3%
αααα	15	3%
παο (ελληνικά ή λατινικά)	13	3%
$#@!	8	2%
θεος	3	1%

Ερώτηση 2:

Ποιοι από τους παρακάτω κωδικούς πιστεύετε πως είναι κοινοί/μη-ασφαλείς;

Πίνακας 5.6: Δεδομένα δεύτερης ερώτησης.

Ποσοστό κωδικών ως κοινοί/μη ασφαλής		
Επιλογές	**Απαντήσεις**	**Ποσοστό**
123456	512	86%
password	342	58%
qwertyui	233	39%
@$fale!@	37	6%
mercedes	245	41%
Ak%T2h	20	3%
Batman	261	44%
F!reB@!!	42	7%
xxxxxxxx	319	54%

Ερώτηση 3:

Πόσο συχνά αλλάζετε τους κωδικούς σας πρόσβασης;

Πίνακας 5.7: Δεδομένα τρίτης ερώτησης.

Συχνότητα αλλαγής κωδικών από τους χρήστες		
Επιλογές	**Απαντήσεις**	**Ποσοστό**
Ποτέ	233	39%
Κάθε χρόνο	168	28%
Κάθε 6 μήνες	84	14%
Πολύ συχνά	78	13%
Τυχαία/ όποτε υπάρχει πρόβλημα	31	5%

Ερώτηση 4:

Χρησιμοποιείτε στα password σας σύμβολα, αριθμούς ή σημεία στίξης;

Πίνακας 5.8: Δεδομένα τέταρτης ερώτησης.

Χρήση Πολυπλοκότητας		
Επιλογές	**Απαντήσεις**	**Ποσοστό**
ΝΑΙ	471	79%
ΟΧΙ	123	21%

Ερώτηση 5:

Πόσους συνολικά κωδικούς πρόσβασης χρησιμοποιείτε για ιστοσελίδες στο internet;

Πίνακας 5.9: Δεδομένα πέμπτης ερώτησης.

Πλήθος κωδικών που χρησιμοποιούν οι χρήστες		
Επιλογές	**Απαντήσεις**	**Ποσοστό**
Μόνο ένα	188	32%
Δύο με τρείς	287	48%
Διαφορετικό για κάθε ιστοσελίδα	101	17%
Άλλο	18	3%

Ερώτηση 6:

Πόσοι χαρακτήρες είναι αρκετοί για να προσδώσουν ασφάλεια σε έναν κωδικό;

Πίνακας 5.10: Δεδομένα τέταρτης ερώτησης.

Αριθμός χαρακτήρων κωδικού πρόσβασης		
Επιλογές	**Απαντήσεις**	**Ποσοστό**
Τέσσερις με πέντε	48	8%
Έξι με εφτά	145	24%
Οχτώ με δέκα	273	46%
Περισσότεροι από δέκα	128	22%

Ερώτηση 7:

Από πόσους χαρακτήρες αποτελούνται συνήθως οι δικοί σας κωδικοί πρόσβασης;

Πίνακας 5.11: Δεδομένα έβδομης ερώτησης.

Αριθμός χαρακτήρων κωδικού πρόσβασης		
Επιλογές	**Απαντήσεις**	**Ποσοστό**
Τέσσερις με πέντε	38	6%
Έξι με εφτά	146	25%
Οχτώ με δέκα	285	48%
Περισσότεροι από δέκα	125	21%

Ερώτηση 8:

Ποιόν τρόπο χρησιμοποιείτε για να θυμάστε τον κωδικό σας;

Πίνακας 5.12: Δεδομένα όγδοης ερώτησης.

Τρόποι αποθήκευσης των κωδικών		
Επιλογές	**Απαντήσεις**	**Ποσοστό**
Απομνημόνευση	471	79%
Αποθήκευση στον browser (internet explorer, Firefox κτλ.)	110	19%
Αποθήκευση χειρόγραφα	102	17%
Αποθήκευση τοπικά στον Η/Υ σας	84	14%
Άλλο	18	3%

Ερώτηση 9:

Χρησιμοποιείτε κάποια από τα παρακάτω προσωπικά σας στοιχεία ως κωδικό πρόσβασης;

Πίνακας 5.13: Δεδομένα ένατης ερώτησης.

Ποσοστά προσωπικών στοιχείων που αποτελούν κωδικούς πρόσβασης.		
Επιλογές	**Απαντήσεις**	**Ποσοστό**
Ημερομηνία γέννησης	183	31%
Το όνομά μου	154	26%
Αθλητική ομάδα	85	14%
Κινητό τηλέφωνο	58	10%
Άλλο	60	10%
Διεύθυνση	51	9%
Μάρκα/μοντέλο αυτοκινήτου	46	8%
Επωνυμία εταιρείας που εργάζομαι	35	6%
Όνομα κατοικίδιου ζώου	31	5%
Όνομα συζύγου	30	5%
Όνομα παιδιών	29	5%
Πολιτικό κόμμα	5	1%
Κανένα	166	28%

Ερώτηση 10:

Έχετε αποκαλύψει ποτέ τον κωδικό σας σε κάποιον γνωστό σας για να το χρησιμοποιήσει εκ μέρους σας;

Πίνακας 5.14: Δεδομένα δέκατης ερώτησης.

Αποκάλυψη του κωδικού		
Επιλογές	**Απαντήσεις**	**Ποσοστό**
ΝΑΙ	271	46%
ΟΧΙ	323	54%

Ερώτηση 11:

Έχετε αποκαλύψει κάποιον από τους κωδικούς σας, στον/ην σύζυγο/σύντροφο σας;

Πίνακας 5.15: Δεδομένα ενδέκατης ερώτησης.

Αποκάλυψη του κωδικού σε σύζυγο/σύντροφο		
Επιλογές	**Απαντήσεις**	**Ποσοστό**
ΝΑΙ	200	66%
ΟΧΙ	394	34%

Ερώτηση 12:

Πιστεύετε πως έχει ποτέ παραβιαστεί ο κωδικός πρόσβασής σας σε κάποια εφαρμογή;

Πίνακας 5.16: Δεδομένα δωδέκατης ερώτησης.

Πιθανή παραβίαση του κωδικού		
Επιλογές	**Απαντήσεις**	**Ποσοστό**
ΝΑΙ	171	29%
ΟΧΙ	423	71%

Ερώτηση 13:

Ποιοί από τους παρακάτω κωδικούς πρόσβασης που χρησιμοποιούνται από έλληνες χρήστες θεωρείται πως είναι κοινοί/μη-ασφαλής;

Πίνακας 5.17: Δεδομένα δέκατης τρίτης ερώτησης.

Αδύναμοι κωδικοί από Έλληνες χρήστες.		
Επιλογές	**Εμφανίζεις**	**Ποσοστό**
Το όνομά μου	451	76%
Αθλητική ομάδα	412	69%
Όνομα πόλης	329	55%
Το επίθετό μου	329	55%
Όνομα αλυσίδας καταστημάτων	152	26%
Όνομα κατοικίδιου ζώου	147	25%
Αριθμός πινακίδας αυτοκινήτου	129	22%
Άλλο	21	4%

ΕΥΡΕΤΗΡΙΟ ΟΡΩΝ